



## **Bloque 4. MUESTREO Y TRATAMIENTO DE DATOS (3 sesiones)**

### **OBJETIVOS**

- ✓ Presentar los problemas y métodos relacionados con la obtención y el tratamiento de datos cuantitativos en ecología.
- ✓ Plantear y contrastar hipótesis sobre relaciones y procesos ecológicos.
- ✓ Conocer, en particular, métodos para la investigación de la estructura de las comunidades vegetales en relación con algún factor físico: muestreo (sesión 1), tratamiento estadístico de datos (sesiones 2 y 3).

### **PLANTEAMIENTO DEL PROBLEMA**

Se pretende investigar la posible influencia del tipo de sustrato (ácido/básico) sobre la diversidad y la abundancia de fanerófitos y caméfitos leñosos y alguna herbácea perenne (ver Anexo 4.1). Para ello, en la primera sesión se pondrán en práctica técnicas de muestreo y se tomarán los datos que serán analizados en sesiones posteriores.

### **Sesión 1 (campo) MUESTREO DE VEGETACIÓN**

- ✓ El área de estudio se ubica en los mismos lugares de El Vellón (Madrid) objeto de las prácticas de cartografía y definición de unidades ambientales. Se conoce por tanto la existencia de diferentes sustratos líticos en el área, cuya influencia en la vegetación es lo que se quiere investigar (ver Anexo 4.2 para una síntesis de la flora en El Vellón).
- ✓ Para ello se diseña un muestreo (ver Anexo 4.1) recogiendo las dos situaciones posibles: sustrato ácido (gneises) y sustrato básico (Cretácico), muestreo que será completado por la totalidad de los alumnos matriculados.

### **MATERIAL:**

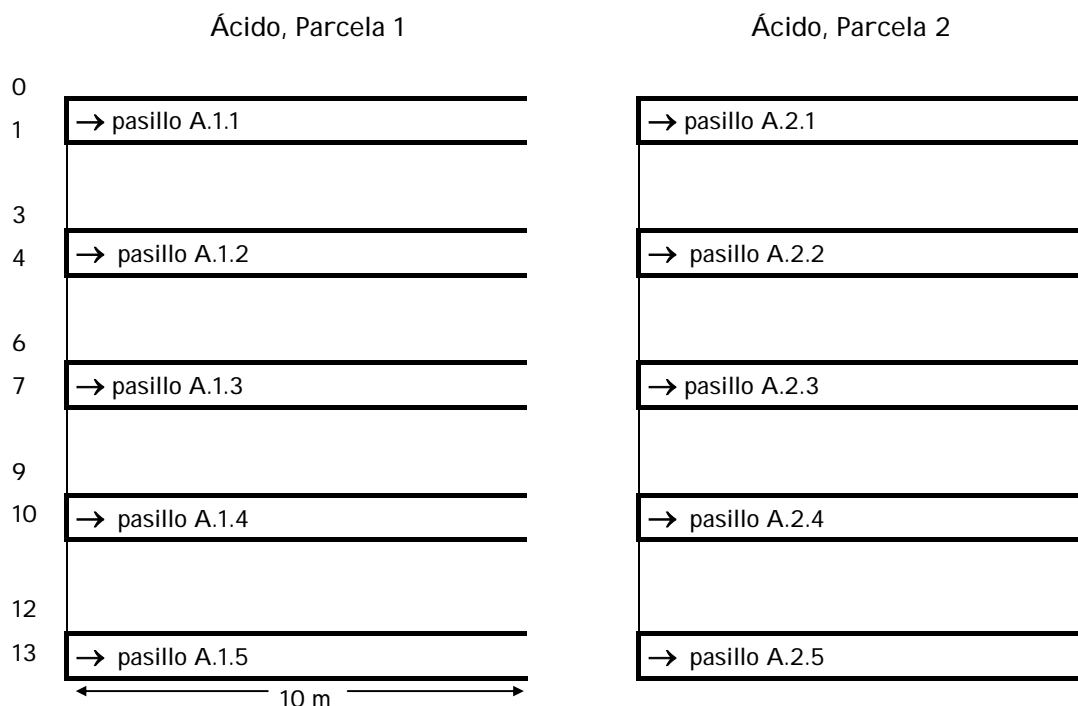
- |                    |                               |
|--------------------|-------------------------------|
| - Cintas métricas. | - Palos de 1 m.               |
| - Clinómetros.     | - Altimetros.                 |
| - Brújulas.        | - Estadillos (ver Anexo 4.3). |

### **PROCEDIMIENTO:**

1. Se quiere medir la cobertura lineal de distintas especies de plantas en uno y otro sector. En cada uno de ellos, ácido o básico, se prevé estudiar, a lo largo de todos los días de muestreo, hasta un total de 80 unidades de muestreo, que serán pasillos de 10 x 1 m. Para ello y con ayuda de cintas métricas, cada día se dispondrán al azar en cada sustrato 2 ejes longitudinales (parcelas 1 y 2) en el sentido de la pendiente de la ladera, a lo largo de los



cuales, y transversalmente, se situarán de modo regular 5 pasillos por parcela, separados en distancias de 2 m., según el siguiente esquema:



2. Una vez en el campo y situados los pasillos, y por razones de organización, éstos se numerarán antes de comenzar el trabajo, según el ejemplo. A cada grupo de 5 alumnos se le asignará un número de pasillo. Una vez realizado el muestreo en uno de los sustratos, se trasladarán al otro tipo de terreno y estudiarán el pasillo equivalente (ej: A.1.1. en ácido y B.1.1. en básico)
3. Se muestreará la variable cobertura lineal en el sentido del pasillo de las distintas especies de fanerófitos y caméfitos presentes, anotándose la medida de cada especie por pasillo en el estadillo adjunto (Anexo 4.3). Además, deberán cumplimentarse los siguientes datos:
  - Localidad y fecha del muestreo.
  - Nombres de los integrantes del grupo de muestreo.
  - Altitud, pendiente y orientación de cada parcela, así como la pendiente transversal de cada pasillo, medida a su inicio.
4. Las especies que aparecen en el estadillo son las que con mayor probabilidad encontraremos (ver Anexo 4.2). No obstante si aparece alguna no incluida se añadirá al final del estadillo. Una vez en la zona se aprenderá a distinguirlas todas.
5. Es imprescindible una absoluta pulcritud, tanto en la toma de datos como en su consignación en los estadillos. Aprovechad el espacio disponible para escribir los datos; en el caso de que para alguna especie no quepan las coberturas parciales, continuad el registro ordenadamente al final del estadillo, indicando de qué especie y pasillo se trata.



6. Los estadillos completados de cada grupo (un pasillo de ácido y otro de básico), incluidas las sumas de los totales, serán entregados al profesor, que compondrá la matriz de todo el muestreo para las sesiones de tratamiento de datos.

**BIBLIOGRAFÍA:** Esta bibliografía se aplica a todas las sesiones de este bloque 4.

- Calvo-Sendin J.F. 1994. *Ecología general. Prácticas y experiencias*. Universidad de Murcia.
- Fowler, J. y Cohen, L. 1999. *Estadística básica en ornitología*. SEO-BirdLife. Madrid.
- Spiegel, R. 2001. *Estadística*. McGraw-Hill. Madrid.
- Montes, C.; Ramírez-Díaz, L. 1978. *Descripción y muestreo de poblaciones y comunidades vegetales y animales*. Publicación de la Universidad de Sevilla. Sevilla
- Parker, R.E. 1981 *Estadística para biólogos*. Cuadernos de Biología. Madrid.
- Sokal, R.R. y Rohlf, F.J. 1979 *Biometría. Principios y métodos estadísticos en la investigación biológica*. Ed. Blume. Madrid.

---

#### Anexo 4.1. DESCRIPCIÓN DE LA VEGETACIÓN

Pueden plantearse dos tipos de aproximaciones: la fisionómica y el análisis florístico. Con cualquiera de ellas pueden aplicarse técnicas de muestreo.

##### Fisionomía

Es una combinación de caracteres:

- Estructurales: en relación con la estructura horizontal (mosaico de bosques y prados, espacio entre árboles etc.), y, sobre todo, vertical de la vegetación. Se trata de reconocer diferentes estratos y expresarlo en forma de porcentaje. También la representación con un perfil es muy apropiada (esquema de un bosque, pisos bioclimáticos, orlas de vegetación alrededor de una laguna, etc).
- Funcionales se pueden referir al color, la textura, la fenología, la periodicidad (ej.: bosques caducifolio y perennifolio), la vitalidad o buen desarrollo de las plantas, entre otros.

Pueden servir de ejemplo los tipos biológicos: una clasificación de las plantas superiores basada en la posición de sus yemas respecto al suelo en la estación desfavorable de su ciclo biológico (invierno o, en clima mediterráneo, frecuentemente verano):

- *FANEROFITAS*: yemas a, por lo menos, 25 cm del suelo (árboles, arbustos y lianas).
- *CAMEFITAS*: yemas a menos de 25 cm del suelo (suculentas, subfruticasas).



- *HEMICRIPTOFITAS*: yemas en la superficie del suelo (rosetas, especies con estolones, gramíneas perennes, etc.).
- *CRIPTOFITAS*: yemas bajo tierra o subacuáticas (geófitas, hidrófitas, helófitas).
- *TEROFITAS*: plantas anuales que pasan el verano en forma de semillas.

### Análisis florístico

Hay dos grandes estrategias, una implica la destrucción del material estudiado; la otra se basa en su conservación. Como medidas no destructivas, se pueden definir:

- La densidad: número de individuos por unidad de superficie.
- La cobertura: muy utilizada en comunidades compuestas de muchas especies pequeñas, especies con estolones o con talos, donde no se puede diferenciar un individuo del otro. Se define como el porcentaje de suelo ocupado por la proyección ortogonal de las partes aéreas de las especies vegetales. Existen varios métodos, de los cuales se destacan los 3 siguientes:
  - Escala de Braun-Blanquet: el tanto por ciento de la superficie de la unidad de muestreo que es recubierta por la proyección vertical de las partes aéreas de las especies, en clases (datos semi-cuantitativos).

+: menos de 1%	1: de 1 a 5%
2: de 6 a 25%	3: de 26 a 50%
4: de 51 a 75%	5: de 76 a 100%
  - Cobertura puntual: se hace deslizar una aguja verticalmente y se anotan todos los contactos entre aguja y la vegetación.
  - Cobertura lineal: se establece como la intercepción lineal de la proyección vertical de las partes aéreas de las especies sobre una cinta métrica, a lo largo de un pasillo de ancho y longitud fijadas de antemano. Este es el método que nosotros usaremos.
- La frecuencia: Número de unidades de muestreo en que aparece una especie por el número total de unidades.

### El muestreo

Para conseguir más detalle y poder aplicar técnicas de tratamientos de datos es necesario efectuar medidas sobre una parte representativa de la vegetación. Esto se consigue mediante los muestreos, en los que es importante maximizar la relación entre información recogida y tiempo invertido. Pueden mencionarse 4 tipos básicos de muestreo:

- *REPRESENTATIVO* o subjetivo: Se estudian las unidades de vegetación situadas en las áreas que tienen alguna característica especial. Es muy empleado en fitosociología. En ciertas circunstancias, por ejemplo en zonas de acceso difícil, es la única posibilidad. Los



datos recogidos con este muestreo no son siempre apropiados para realizar pruebas estadísticas.

- *ALEATORIO*: Cada punto del territorio tiene igual probabilidad de ser elegido. Por lo tanto, es inapropiado si la zona es heterogénea, sobre todo cuando existe un mosaico de formaciones de diferentes superficies (bosques con pequeños claros, setos y prado).
- *REGULAR* o sistemático: Se estudian las unidades de muestreo (parcelas) dispuestas en arreglo regular, cuadrículas o líneas. La forma más utilizada en este tipo de muestreo es el transecto donde las unidades de muestreo están situadas en bandas, contiguas o no (interesante cuando existe un gradiente de altitud, de humedad, etc.).
- *ESTRATIFICADO*: Se subdivide el territorio en partes homogéneas y dentro de cada una se muestrea al azar o sistemáticamente.

Una vez elegido el tipo de muestreo el siguiente paso es abordar su diseño. El diseño del muestreo es una tarea muy importante y muy delicada. Se establece según el fin propuesto, el problema que se investigue, las hipótesis de partida etc. Hace falta determinar:

- La forma y la superficie de la unidad de muestreo: Pueden utilizarse círculos, cuadrados o pasillos, pero la forma elegida tiene que mantenerse para todo el muestreo. La superficie depende del tamaño de las especies. El muestreo de una pradera con especies herbáceas requiere, por ejemplo, un cuadro de 20 a 50 cm de lado. Si se trata de caméfitos posiblemente sea necesario aumentar la superficie de 1 a 5 m<sup>2</sup>. En el caso de matorral alto se eligen unidades de 10 a 100 m<sup>2</sup>.
- El número de unidades de muestreo: Está en relación con el tiempo a invertir, el detalle requerido y unas consideraciones estadísticas. Muchas veces, sin embargo, prevalece la experiencia.
- Los puntos al azar: Existen varios métodos para determinarlos, cada uno con sus ventajas y desventajas. Uno sencillo consiste en situar sobre el terreno, el mapa o la foto aérea ejes de coordenadas procediendo a continuación a extraer parejas de coordenadas (tantas como puntos se requieran) en una tabla de números al azar o con un programa de ordenador.

Por último, es necesario tomar ciertos datos cualitativos. Es imprescindible anotar la fecha del muestro y la localidad donde se ha efectuado. Deben hacerse medidas de altitud, orientación, pendiente, etc., de cada unidad de muestreo. Otras observaciones pueden ser de gran utilidad para encontrar una explicación de algunos resultados: pedregosidad, característica edáficas, intervenciones humanas, etc. Estos datos pueden convertirse en datos semi-cuantitativos.

En El Vellón, nosotros adoptaremos una aproximación de análisis florístico, mediante un muestreo estratificado según naturaleza del sustrato (ácido/básico) y posterior disposición aleatoria de las parcelas. En cada parcela, las unidades de muestreo (pasillos de 1 m. de ancho y 10 m. de largo) se dispondrán de modo regular separadas 2 m. La variable objeto de medida



será la cobertura lineal de cada especie en el sentido del pasillo, y las unidades de medida serán centímetros.

---

## Anexo 4.2. VEGETACIÓN Y FLORA DE EL VELLÓN

### ENCINARES MEDITERRÁNEOS

Potencialmente, bosque de encina (*Quercus ilex* subsp. *ballota*) en clima mediterráneo continentalizado, con abundante enebro de miera (*Juniperus oxycedrus*).

#### Flora del encinar carpetano

Encinares sobre sustratos ácidos (gneises). Frecuentemente aclarados para producción de ganadería bovina, resultando un paisaje de pastizales con cercas de piedra perimetrales y las encinas y otros elementos florísticos a modo de setos lineales. El declive reciente de la actividad ganadera genera distintos estadios de infra-pastoreo, resultando patrones/unidades de a) Prados con cercas arboladas, b) Arbolado disperso-matorral, y c) Arbolado denso forestal. Entre otras especies se encuentran:

- Cantueso (*Lavandula pedunculata*)
- Falsa jarilla (*Halimium umbellatum* subsp. *viscosum*)
- Tomillo (*Thymus mastichina*)
- Retama negra (*Cytisus scoparius*)
- Botonera (*Santolina rosmarinifolia*)
- Retama de bolas (*Retama sphaerocarpa*)
- Aladierno (*Rhamnus alaternus*)
- Olivilla (*Phyllirea angustifolia*)
- Torvisco (*Daphne gnidium*)
- Rubia (*Rubia peregrina*)
- Esparraguera (*Asparagus acutifolius*)
- Madreselva (*Lonicera periclymenum*)

#### Flora del encinar manchego

Encinares sobre sustratos básicos (Cretácico y Terciario). Frecuentemente sustituido por cultivos de cereal y de olivo y talado para abrir pastos aprovechados por ovino y caprino. Como resultado aparecen formaciones de sustitución como aulagares, tomillares, espartales esplegares y sus mezclas. El declive reciente de la actividad ganadera genera distintos estadios de abandono, resultando patrones/unidades de a) Arbolado denso-forestal, b) Arbolado disperso-matorral, c) Cultivos abandonados-matorral, d) Secano extensivo, y e) Secano intensivo. Entre otras especies se encuentran:

- Espliego (*Lavandula latifolia*)
- Aulaga (*Genista scorpius*)
- Esparto (*Stipa tenacissima*)
- Tomillo salsero (*Thymus vulgaris*)



- Lino (*Linum suffruticosum*)
- Lastón (*Brachypodium retusum*)
- Flomis (*Phlomis lychnitis*)
- Estaelina (*Staehelina dubia*)
- Tomillo (*Thymus zygis*)
- Aladierno (*Rhamnus alaternus*)
- Retama de bolas (*Retama sphaerocarpa*)
- Aladierno o espino negro (*Rhamnus lyciodes*)
- Torvisco (*Daphne gnidium*)
- Esparraguera (*Asparagus acutifolius*)

### Flora de los fondos de barrancos, umbrías y zonas de descarga

Sobre sustratos básicos, en ubicaciones con menor evapotranspiración, suelos más profundos y humedad edáfica en verano, aparecen quejigos (*Quercus faginea* subsp. *faginea*) y formaciones freatofíticas, que vimos acompañados de:

- Arce de Montpellier (*Acer monspessulanum*)
- Cornicabra (*Pistacia terebinthus*)
- Olmo (*Ulmus minor*) También en riberas.
- Chopo (*Populus nigra*) También en riberas.
- Majuelo (*Crataegus monogyna*)
- Jasmín (*Jasminum fruticans*)
- Junco churrero (*Scirpus holoschoenus*)
- Aligustre (*Ligustrum vulgare*)
- Escaramujo o Rosal silvestre (*Rosa* spp.)
- Zarza (*Rubus* spp.)
- Aladierno espino negro (*Rhamnus lyciodes*)
- Esparraguera (*Asparagus acutifolius*)

### Flora de las riberas

El bosque en galería original se encuentra muy alterado por la actividad humana que prácticamente lo ha eliminado de las terrazas fluviales para instalar huertas y regadíos. Jalonando cauces y lechos de inundación, aún se encuentran, entre otras:

- Sauces (*Salix* spp.)
- Chopo (*Populus nigra*)
- Olmo (*Ulmus minor*)
- Ailanto (*Ailanthus altissima*)
- Escaramujo o Rosal silvestre (*Rosa* spp.)
- Zarza (*Rubus* spp.)
- Junco churrero (*Scirpus holoschoenus*)



Anexo 4.3. PLANTILLA DE MUESTREO

**Localidad:**

**Coordenadas UTM:**

**Fecha:**

**Pendiente:**

**Orientación:**

**Grupo de prácticas:**

**Nombres de los integrantes del grupo:**

ESPECIE	PASILLO Nº	COBERTURAS PARCIALES	COBERTURA TOTAL
<i>Argyrolobium zanonii</i>			
<i>Artemisia campestris</i>			
<i>Cytisus scoparius</i>			
<i>Daphne gnidium</i>			
<i>Fumana ericoides</i>			
<i>Genista scorpius</i>			
<i>Halimium umbellatum</i> <i>subsp. viscosum</i>			
<i>Helianthemum apenninum</i>			
<i>Helianthemum cinereum</i>			
<i>Helichrysum stoechas</i>			
<i>Juniperus oxycedrus</i>			
<i>Lavandula latifolia</i>			
<i>Lavandula stoechas</i> <i>subsp. pedunculata</i>			
<i>Linum suffruticosum</i>			
<i>Phlomis lychnitis</i>			
<i>Quercus (ilex subsp.)</i> <i>ballota</i>			
<i>Retama sphaerocarpa</i>			
<i>Ruta montana</i>			
<i>Santolina rosmarinifolia</i>			
<i>Staehelina dubia</i>			
<i>Teucrium polium</i>			
<i>Thymus mastichina</i>			
<i>Thymus vulgaris</i>			
<i>Thymus zygis</i>			



## Sesión 2 (Aulas de informática) TRATAMIENTO DE DATOS I: Estadística básica

- Se pretende conocer y ejercitarse el manejo de algunas técnicas básicas de análisis e interpretación de datos cuantitativos, con ayuda de diversos métodos estadísticos y numéricos.
- Para ello se trabajará sobre los datos obtenidos en la sesión anterior de muestreo de la vegetación de El Vellón, referidos a la composición y la estructura de las comunidades de plantas leñosas sobre sustratos ácido y básico. Sus posibles similitudes y diferencias serán objeto de distintas cuestiones a las que se tratará de responder en esta y la siguiente sesión de prácticas.

### MATERIAL:

- Matriz de datos del muestreo.
- Hoja de cálculo Excel.
- Tablas estadísticas.
- Papel milimetrado.

### PROCEDIMIENTO:

Al empezar, es aconsejable hacer un repaso de conceptos generales sobre el muestreo, recapitulando la técnica empleada en nuestro caso. También se recordarán los conceptos de población, muestra, unidad de muestreo, tamaño muestral y variable (ver Anexo 4.4).

1. El primer paso consistirá en comprobar si el esfuerzo de muestreo realizado ha sido suficiente para cubrir un área mínima en la que esté presente la variedad de especies de plantas leñosas existentes en uno y otro sectores muestreados.

Para ello se representará gráficamente, sobre papel milimetrado, la evolución del número de especies acumulado a medida que se aumenta el área muestreada. La inspección de la curva así obtenida permitirá interpretar si se ha alcanzado el área mínima de muestreo para recoger adecuadamente la variedad de la vegetación estudiada.

- Una curva en permanente incremento nos hará suponer que de haber aumentado el área estudiada habríamos encontrado nuevas especies, que faltan en nuestros datos.
- Una curva que tiende a estabilizarse indicará que hemos encontrado la práctica totalidad de las especies que nos interesaban.

2. A continuación se procederá a la comparación de las listas de especies, que podemos agrupar en especies presentes sobre sustrato ácido y especies presentes sobre sustrato básico. A partir de ahí hay que separar:

- Especies exclusivas del sustrato ácido.
- Especies exclusivas del sustrato básico.
- Especies comunes a ambos sustratos.



Las que sean exclusivas de uno u otro sector podrían considerarse tentativamente como indicadoras de la dominancia de uno u otro factor. Sobre aquellas otras que aparecen en ambos tipos de sustrato de momento no podemos obtener conclusiones.

3. ¿Qué sector presenta una mayor riqueza de especies, es decir, un mayor número de especies distintas? ¿Y cuál de los dos presenta una mayor diversidad?

La contestación de la primera pregunta será inmediata. Para responder a la segunda, sin embargo, habremos de introducir el concepto de diversidad, que tiene en cuenta no sólo la cantidad de especies sino cómo son de abundantes sus respectivos individuos.

Para obtener un índice cuantitativo y comparable de esta propiedad de las comunidades vivientes son varios los métodos propuestos. Uno de los más utilizados es el índice de diversidad de Shannon-Wiener, que para  $n$  especies se calcula como:

$$H = -\sum_{i=1}^n p_i \cdot \log_2 p_i \quad \text{donde:} \quad \log_2 p_i = \frac{\log p_i}{\log 2}$$

siendo  $p_i$  la proporción de la especie  $i$  respecto al total (cobertura  $i$  / cobertura total).

$H$  se expresa en bits de información, ya que el índice de Shannon-Wiener es, en realidad, una medida de la cantidad de información contenida en cualquier tipo de sistema.

4. Tras este somero análisis de algunas características generales de la vegetación de los dos sectores, quedará la cuestión de interpretar el significado de aquellas especies que sean comunes a ambos. Su mera aparición en los dos sustratos no significa que sean indiferentes a los factores ecológicos que actúan en uno u otro lugar, ya que su abundancia puede ser muy diferente. Para tratar este problema se van a aplicar algunas técnicas de estadística descriptiva. Para los datos de cobertura de cada una de las especies, tanto en sustrato ácido como básico, calcularemos los siguientes valores, cuyo significado se comentará brevemente:

- La media: 
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- La varianza: 
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- La desviación típica: 
$$s = \sqrt{s^2}$$

- El coeficiente de variación: 
$$CV = \frac{s}{|\bar{x}|} \quad \text{o bien (expresado en \%)} \quad CV = \left( \frac{s}{|\bar{x}|} \right) 100$$



5. El siguiente paso es comparar las medias de cobertura de cada especie en uno y otro sector. Pero antes convendrá recordar que nuestras medias muestrales no son más que estimas de la media poblacional, un parámetro cuyo valor exacto no podemos conocer. Podemos sin embargo evaluar qué grado de precisión tienen nuestras estimas, lo cual dependerá del tamaño de la muestra y de la variabilidad de la población.

Para ello calcularemos para cada media un intervalo de confianza, que establecerá el rango de valores por encima y por debajo en el que, con una determinada probabilidad, por ejemplo de 0.95 (es decir, del 95%), se sitúa la media poblacional. El razonamiento matemático (Teorema del Límite Central) se basa en que, para poblaciones con una distribución normal, puede demostrarse que las medias muestrales que obtendríamos si repitiéramos infinitas veces la extracción de una muestra siguen también, a su vez, una distribución normal cuya media es la media poblacional y cuya desviación típica es la desviación típica de las medias muestrales, también llamado error típico de la media (ET):

$$E.T. = \frac{s}{\sqrt{n}} \quad \text{donde: } s: \text{ desviación típica de la muestra y } n: \text{ número de observaciones}$$

En tales condiciones, el intervalo de confianza para una media puede calcularse con una probabilidad  $1 - \alpha = 0.95$  (por tanto,  $\alpha = 0.05$ ) como:

$$\bar{x} \pm t_{0.05, n-1} \cdot \frac{s}{\sqrt{n}}$$

donde  $n-1$  son los grados de libertad y el valor buscado de  $t$  puede consultarse en una tabla de distribución de la  $t$  de Student de dos colas (Anexo 4.5). Es fácil advertir que puede conseguirse una mayor precisión, es decir, un intervalo de confianza más pequeño, aumentando el esfuerzo de muestreo, es decir, aumentando  $n$ . Inversamente, una mayor variabilidad de la población, reflejada en una  $s$  mayor, disminuirá la precisión y hará mayor el intervalo.

---

#### Anexo 4.4. ESTADÍSTICA BÁSICA Y DESCRIPTIVA

Como es propio de las disciplinas científicas, la ecología intenta conocer e interpretar el mundo que nos rodea. En particular, la ecología se ocupa del mundo natural y de las relaciones que mantienen sus componentes: es la *historia natural científica* (Elton, 1927).

En su empeño por conocer la naturaleza viviente, los ecólogos se ven frecuentemente en la tesitura de acumular datos con los cuales discernir patrones, para deducir a partir de ellos los procesos subyacentes al funcionamiento de la naturaleza. La *estadística* es la herramienta de la que se sirve el ecólogo para explorar y analizar los datos que recoge de la naturaleza. Aunque no es sino un medio, es lo suficientemente importante para que sea preciso conocerlo adecuadamente si deseamos que cumpla su misión.

En primer lugar, es necesario hacer una distinción entre *estadística descriptiva* y *estadística inferencial*. La primera se refiere a aquellos procedimientos que tienen por objeto describir el conjunto de observaciones que hemos realizado. ¿Cuál es el contenido medio en nutrientes de los embalses españoles? ¿Cómo de variable es el tamaño de puesta de un petirrojo a lo largo



de su vida? La inferencia estadística, como su propio nombre indica, pretende obtener inferencias (predicciones) sobre atributos de un conjunto de objetos (población) a partir de la observación de un subconjunto de aquél (muestra). ¿Difieren dos poblaciones de jabalíes en su tamaño medio? ¿Existe relación entre el tamaño de un depredador y el de las presas que consume?

El valor de una característica de la población (por ejemplo, número medio de parásitos que infestan un hospedador en una población natural de nemátodos) se denomina parámetro. Con frecuencia, sin embargo, ese valor no se conoce pues resulta difícil, si no imposible, estudiar el total de la población. La medida de tal característica en una muestra de la población permite hacer inferencias sobre el valor del parámetro, recibiendo el nombre de estadístico. Existen técnicas estadísticas que trabajan sobre los parámetros de una población y reciben por ello el nombre de *estadística paramétrica*. Parece razonable que las que no lo hacen así se denominen *estadística no-paramétrica*.

Es fácil imaginar que no resulta fácil inferir características fiables de la población a partir de una cantidad limitada de observaciones (muestras). Para cumplir satisfactoriamente su misión, los métodos estadísticos se ayudan ante esa dificultad de una serie de asunciones o suposiciones sobre la estructura o distribución estadística de los datos bajo estudio.

Se ha comprobado que muchos atributos de las poblaciones naturales tienen una distribución normal (nombre que recibe esa distribución precisamente por ser extraordinariamente común) y, por lo tanto se han desarrollado técnicas estadísticas que funcionan expresamente en aquellos casos en los que la variable de trabajo tiene esa distribución en la población. Técnicas de este tipo se denominan no independientes de la distribución. Aquellos métodos que no hacen ningún tipo de suposición sobre la distribución de la variable estudiada se llaman independientes de la distribución. En un intento de simplificar la terminología (y de confundir los significados) se puede encontrar con frecuencia que al hablar de estadística no-paramétrica los autores se refieren tanto a ésta, en sentido estricto, como a la estadística independiente de la distribución (o de distribución libre).

Aquella propiedad que difiere de unos individuos a otros y que nosotros registramos se denomina variable (altura, conductividad, abundancia, color...). Las variables pueden ser de muy diversos tipos lo que condiciona necesariamente la técnica estadística que podremos emplear con ellas. Una clasificación de las variables según su naturaleza puede ser la siguiente:

- Variables cuantitativas:
  - Continuas (altura de una planta, biomasa de un insecto...).
  - Discontinuas, merísticas o discretas (abundancia de una especie, número de escamas de la línea lateral de un pez...).
- Variables cualitativas:
  - Ordinales (por ejemplo, rangos de abundancia: abundante, frecuente, raro).
  - Nominales (color, sexo, localidad...).



Obviamente, diferentes tipos de variables han de tratarse con diferentes técnicas. De la misma forma, diferentes preguntas pueden ser respondidas empleando distintos métodos.

Anexo 4.5. TABLA DE LA DISTRIBUCION  $t$ -Student

**Tabla I**  $t$  de Student. Valores excedidos con probabilidad  $p$

Grados de libertad	$p = 0,1$	0,05	0,02	0,01	0,002	0,001
1	6,314	12,706	31,821	63,657	318,31	636,62
2	2,920	4,303	6,965	9,925	22,327	31,598
3	2,353	3,182	4,541	5,841	10,214	12,924
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,893	6,869
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,610	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,767
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,690
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,659
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
60	1,671	2,000	2,390	2,660	3,232	3,460
120	1,658	1,980	2,358	2,617	3,160	3,373
$\infty$	1,645	1,960	2,326	2,576	3,090	3,291



### Sesión 3 (Aulas de informática). TRATAMIENTO DE DATOS II: Contraste de hipótesis

Tras haber sometido a una primera elaboración a los datos procedentes del muestreo, en esta sesión se planteará la cuestión de qué métodos estadísticos podemos emplear para llegar a conclusiones fiables sobre si la presencia de una especie muestra diferencias significativas entre uno y otro sustrato. Para ello convendrá recordar los conceptos pertinentes a toda una serie de pruebas estadísticas conocidas como contrastes de hipótesis (Anexo 4.6).

#### PROCEDIMIENTO:

1. Una primera aproximación al problema de si la presencia de una especie es igual en el sustrato ácido y en el básico es la de examinar el número de veces que esa especie aparece en los pasillos de vegetación muestreados. En este caso no estaríamos considerando las medidas de cobertura de la especie sino simplemente su presencia o ausencia en cada pasillo. Se trata de una situación frecuente en muestreos y experimentos cuyos resultados consisten en datos que se clasifican en categorías o clases, obteniéndose así frecuencias para cada una de dichas clases. En nuestro caso las unidades muestreadas, los pasillos, se dividirían en dos clases, sustrato ácido o sustrato básico. Por otro lado, el examen de los datos obtenidos permite agruparlas en otras dos clases, presencia de la especie o ausencia de la especie.

Si somos capaces de establecer una hipótesis sobre la distribución esperada de tales observaciones podremos comprobar la significación de una desviación dada entre las frecuencias observadas y las esperadas. Es decir, determinar si la probabilidad de obtener tal desviación debido a efectos aleatorios es probable o no.

El test de la chi cuadrado ( $\chi^2$ ) permite hacer tal tipo de estimaciones. Para ello seguiremos las siguientes etapas:

- a) Formulación de la hipótesis nula  $H_0$ . La distribución de frecuencias observadas se ajusta a una distribución teórica, que nos permite calcular, como luego veremos, unas frecuencias esperadas. En nuestro caso la distribución teórica es que la planta de nuestro interés se distribuya independientemente de la característica ambiental considerada (tipo de sustrato).
- b) Construcción de la tabla de contingencia. Nuestro caso es el más sencillo, el de la tabla de contingencia  $2 \times 2$ , en el cual existen dos clasificaciones, cada una de las cuales se subdivide en dos categorías mutuamente excluyentes. El número de clasificaciones corresponde al número de variables. Por ejemplo, la variable 1 se referiría a la presencia o ausencia de la especie y la variable 2 al tipo de sustrato (ácido o básico).

Si de 100 observaciones (en nuestro caso, pasillos de vegetación muestreados), 40 han sido hechas en ácido y 60 en básico, y si la especie estudiada apareció 30 veces en nuestras parcelas (20 en ácido y 10 en básico) estando ausente en 70 ocasiones (20 en ácido y 50 en básico) se puede realizar la siguiente tabla de contingencia:



Frecuencias observadas

	ÁCIDO	BÁSICO	
PRESENCIA	20 (a)	10 (b)	30 (a+b)
AUSENCIA	20 (c)	50 (d)	70 (c+d)
	40 (a+c)	60 (b+d)	N=100 (a+b+c+d)

- c) Cálculo de las frecuencias esperadas. Dado que si 2 acontecimientos son independientes la probabilidad de que ocurran juntos puede calcularse mediante el producto de sus probabilidades separadas, la frecuencia esperada será, por ejemplo:

Probabilidad de encontrar la especie en sustrato ácido:

$$P_{+,A} = (P_+ \times P_A) \times N = (30/100 \times 40/100) \times 100 = 12$$

Frecuencias esperadas para el caso de distribución de la especie independiente del sustrato

	ÁCIDO	BÁSICO
PRESENCIA	12	18
AUSENCIA	28	42

- d) Cálculo del  $\chi^2$ . La fórmula general es:  $\chi^2 = \sum \frac{(o-e)^2}{e}$  donde  $o$  son las frecuencias observadas y  $e$  las esperadas.

En el caso de una tabla de contingencia de 2 x 2, la fórmula simplificada es

$$\frac{(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)} \times N$$

- e) Rechazo o aceptación de la hipótesis. Se fija un nivel de significación o probabilidad de error (normalmente se utiliza  $\alpha = 0.05$ , es decir que se acepta el riesgo de equivocarse 5 veces sobre 100, si fuese  $\alpha = 0.01$  se aceptaría el riesgo de equivocarse 1 vez sobre 100).

Lectura de la tabla de  $\chi^2$  (ver Anexo 4.7). Para ello es necesario conocer los grados de libertad, que se calculan g.l. = (filas-1) x (columnas-1) de la tabla de contingencia. En este caso el grado de libertad es 1 x 1 = 1. Hay 4 casillas y 3 restricciones. En efecto, se emplean los totales marginales (suma de los valores observados) para calcular las frecuencias esperadas, y esto, como cuando se estime un parámetro produce una pérdida de grados de libertad. Una vez fijado estos totales, una única frecuencia observada insertada en una casilla de la tabla determina todas las demás.

Se toman las probabilidades 0.5 y 0.01, y se lee la intersección con g.l. = 1, es decir 3.84 y 6.64 respectivamente. En la curva 1 el área a la derecha de 3.84 es de 5% y de 6.64, 1%.

Si  $\chi^2$  calculado <  $\chi^2$  teórico, no se rechaza la hipótesis nula ( $H_0$ ). Se puede decir que las frecuencias observadas y esperadas son similares, y por lo tanto que la especie se distribuye independientemente de la variable ambiental considerada.



Si  $\chi^2$  calculado  $>$   $\chi^2$  teórico se rechaza  $H_0$ . Las frecuencias observadas y esperadas son muy diferentes y se puede concluir que la especie prefiere uno de los hábitats (aquél en el cual presentara mayor cobertura).

2. La comparación de la presencia de determinadas especies en el sector ácido y el sector básico puede abordarse de otro modo. Dado que tenemos datos no sólo de la presencia o ausencia de una determinada especie en los pasillos muestreados sino también de su cobertura lineal, disponemos de una medida cuantitativa más precisa de su abundancia en uno y otro caso. Podemos entonces plantearnos la comparación de las medias de cobertura para una especie en uno y otro sustrato.

a. Construimos el nuevo contraste de hipótesis definiendo una hipótesis nula según la cual las medias poblacionales de la especie en ambos sectores son iguales. Estas medias se designan con la letra  $\mu$  y son parámetros que no conocemos. Lo que conocemos son las medias muestrales  $\bar{x}$  que hemos obtenido y que nos proporcionan una estima más o menos fiable. La hipótesis alternativa es, obviamente, que las medias son distintas, es decir, que aunque la especie esté presente en ambos sectores no es indiferente al sustrato, porque presenta abundancias distintas.

$$H_0, \mu_A = \mu_B$$

$$H_1, \mu_A \neq \mu_B$$

b. Suponiendo que ambas poblaciones tienen una distribución normal y que sus varianzas son similares podemos aplicar un test de igualdad de medias basado en la distribución t de Student. Se trata de un test muy sencillo y notablemente robusto, que proporciona resultados bastante fiables aunque los requisitos estadísticos que precisa no se cumplan enteramente. El modo de aplicarlo es el siguiente. Construimos un estadístico t:

$$t_{\text{observado}} = \frac{(\bar{x}_A - \bar{x}_B)}{\sqrt{\left[ \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} \right] \cdot \left( \frac{n_A + n_B}{n_A n_B} \right)}}$$

que, en caso de que  $H_0$  sea cierta, adopta valores cuya distribución probabilística se ajusta a una distribución t de Student. Comprobamos el valor que obtenemos, nuestro t observado, y lo comparamos para un determinado nivel de significación  $\alpha$  (por ejemplo  $\alpha = 0.05$ ) en la tabla de la t de Student (Anexo 4.5) que ya utilizamos cuando calculamos los intervalos de confianza en la sesión anterior. Si nuestro t observado es mayor o menor que unos determinados valores críticos, que son valores teóricos de la t de Student que dejan por encima y por debajo una probabilidad menor que  $\alpha$ , entonces suponemos que  $H_0$  es muy improbable y la rechazamos. Como la diferencia entre ambas medias puede ser positiva o negativa hemos considerado valores de t positivos y negativos y, por tanto, es un test de dos colas.



- c. El valor absoluto, para obviar si es positivo o negativo, del  $t$  observado se compara con un  $t$  crítico que buscamos en la tabla de la distribución de la  $t$  de Student de dos colas, para una probabilidad  $\alpha$ . La distribución de la  $t$  de Student varía según los grados de libertad. En nuestro caso los grados de libertad son  $n_A + n_B - 2$ . Entonces,

si  $|t_{\text{observado}}| < t_{\alpha, n_A + n_B - 2}$ , se acepta  $H_0$ , no hay diferencia entre las medias.

si  $|t_{\text{observado}}| > t_{\alpha, n_A + n_B - 2}$ , se rechaza  $H_0$ , las medias son diferentes.

3. Al comenzar a analizar las diferencias entre la vegetación de uno y otro sustrato ya vimos que podíamos comparar su riqueza y su diversidad de especies, pero entonces lo hicimos para valores globales de una y otra. Ahora, con la técnica de comparación de medias mediante el test de la  $t$  de Student, podemos comparar las respectivas medias de las riquezas y diversidades que se dan en cada pasillo, completando nuestro análisis.

- a. Construimos el nuevo contraste de hipótesis definiendo dos  $H_0$ , según las cuales las riquezas medias de cada sector son iguales y las diversidades medias de cada sector también. Estas medias se designan con la letra  $\mu$  y son, una vez más, parámetros que no conocemos. Lo que conocemos son, las medias muestrales  $\bar{x}$  que hemos obtenido (riqueza media y diversidad media en cada sector). De nuevo, la hipótesis alternativa es que las medias son distintas, es decir, que la riqueza y/o la diversidad de especies no son indiferentes al sustrato, sino que presentan valores distintos en uno y otro.

$$H_0, \mu_A = \mu_B$$

$$H_1, \mu_A \neq \mu_B$$

- b. Suponiendo que ambas poblaciones para cada variable tienen una distribución normal y que sus varianzas son similares podemos aplicar el test de la  $t$  de Student, del mismo modo que en el caso anterior.

Una vez obtenidos los resultados de las distintas comparaciones podemos recapitular y discutir en qué medida la diferencia en el sustrato afecta a la estructura y la composición de la vegetación y, en particular, a la distribución de ciertas especies que hemos estudiado individualmente. En esta discusión convendrá recordar que los dos sectores comparados, ácido y básico, pueden haberse visto afectados, además, por distintos tipos de uso. De este modo podrá discutirse la posibilidad de que alguna de las diferencias detectadas entre sectores este relacionada no solo con las propiedades fisicoquímicas del sustrato sino también con la acción humana que ha favorecido o desfavorecido la presencia de determinadas especies. ¿Podrías diseñar una nueva experiencia para aclarar este tipo de cuestiones?



#### Anexo 4.6. CONTRASTE DE HIPÓTESIS

Aunque diferentes en los supuestos a los que se aplican y en los métodos que emplean, los contrastes de hipótesis coinciden en basarse en la contraposición de dos hipótesis excluyentes. La hipótesis de partida o hipótesis nula, notada como  $H_0$ , es que no pasa nada, que no hay diferencias o efectos en aquello que estamos estudiando. Solo si reunimos evidencias estadísticas suficientes para demostrar que tal hipótesis es muy improbable, pudiendo nosotros fijar ese nivel de improbabilidad, aceptamos la hipótesis alternativa,  $H_A$  o  $H_1$ , que es la que supone la existencia de un efecto o diferencia entre las muestras comparadas.

Llamamos nivel de significación de nuestro contraste a la probabilidad de que nuestra muestra provenga de una población en la que es cierta la hipótesis nula. Si fijamos ese nivel de significación lo que estamos haciendo, por tanto, es decidir qué riesgo estamos dispuestos a asumir de equivocarnos al rechazar una hipótesis nula que es cierta. Dicho de otro modo, cuando logremos rechazar una hipótesis nula, lo haremos admitiendo una cierta probabilidad, que normalmente queremos que sea muy pequeña, de que la hipótesis nula fuera cierta después de todo. Un error de este tipo se llama en estadística error de tipo I y equivale al mencionado nivel de significación (también denominado  $\alpha$ ). Tradicionalmente, se ha empleado como valor estandarizado de  $\alpha$  el del 5%. Quiere esto decir que asumimos una probabilidad de equivocarnos de 1 entre 20 cuando rechazamos la hipótesis nula. O lo que es lo mismo, si rechazamos la hipótesis nula con ese nivel de significación, sólo nos equivocaremos una de cada 20 veces (por término medio), suponiendo que la hipótesis nula es cierta. Por supuesto no existe ninguna regla que impida utilizar otros niveles de significación pero los científicos, al igual que los dueños de los casinos, procuran minimizar sus riesgos de equivocarse y los valores bajos del nivel de significación son siempre apreciados.

El error de tipo I es el que nosotros vamos a fijar y, por tanto, a mantener en unos límites controlados. Conviene saber, sin embargo, que existe otro tipo de error, el error de tipo II, que es la probabilidad de que no rechacemos la hipótesis nula cuando esta sea falsa. Este error tiene que ver con la posibilidad de que aunque exista una diferencia o efecto, es decir, aunque  $H_1$  sea correcta, nuestros datos pueden no ser lo suficientemente concluyentes para demostrarlo estadísticamente.

Es importante señalar que, en puridad, la hipótesis nula nunca se acepta. Si bien en términos coloquiales podemos decir que la aceptamos es importante tener presente que lo que ocurre es que no se tienen evidencias suficientes para poder rechazarla. En muchas ocasiones es suficiente un aumento del tamaño de muestra para que pase a rechazarse la hipótesis que, con menor cantidad de observaciones, no podía caracterizarse como falsa.



Anexo 4.7. TABLA DE LA DISTRIBUCIÓN CHI-CUADRADO

Probabilidad conforme a $H_0$ de que $\chi^2 \geq$ chi cuadrada														
gl	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.00016	.00063	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.29
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	21.62	24.77	27.59	31.00	33.41	40.75
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.24	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70