

Introducción al diseño de experimentos

Luis Cayuela

Mayo de 2011

Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos,
Departamental 1 – DI. 231, c/ Tulipán s/n. E-28933 Móstoles (Madrid),
España. E-mail: luis.cayuela@urjc.es.

Introducción al diseño de experimentos (versión 1.2)

Publicado por: Luis Cayuela



Se autoriza a cualquier persona a utilizar, copiar, distribuir y modificar esta obra con las siguientes condiciones: (1) que se reconozca la autoría de la misma; (2) que no se utilice con fines comerciales; y (3) que si se altera la obra original, el trabajo resultante sea distribuido bajo una licencia similar a ésta.

Para cualquier comentario o sugerencia por favor remitirse al autor de la obra.

Índice

1. Diseños de investigación	4
1.1. El supuesto de independencia	4
1.2. Factores de confusión	5
1.3. Replicación y aleatorización	5
2. Diseño factorial o ANOVA	7
2.1. Supuestos de los diseños factoriales	7
3. Otros diseños de experimentos	9
3.1. Diseño por bloques aleatorizados	9
3.2. Split-plot	11
3.3. ANOVA de medidas repetidas	13
3.4. ANOVA anidado o jerarquizado	16
3.4.1. Pseudorreplicación	16
4. Modelos mixtos	19
4.1. ¿Efectos fijos o aleatorios?	21
4.2. Modelos mixtos ¿una alternativa para representar y analizar otros diseños de experimentos?	21
5. Ejercicios	23
6. Referencias	24

1. Diseños de investigación

Una vez definido el tipo de estudio a realizar y establecer las hipótesis de investigación, el investigador debe concebir la manera práctica y concreta de responder a las preguntas de investigación. Esto implica seleccionar o desarrollar un diseño de investigación y aplicarlo al contexto particular de su estudio. Diseño se refiere al plan o estrategia concebida para responder a las preguntas de investigación. El diseño señala al investigador lo que debe hacer para alcanzar sus objetivos de estudio, contestar las interrogantes que se ha planteado y analizar la certeza de las hipótesis formuladas en un contexto en particular. Hay muchos tipos de diseños distintos: algunos son puramente observaciones y otros manipulativos.

1.1. El supuesto de independencia

La mayoría de los análisis estadísticos asumen que las réplicas son independientes. Independencia en este contexto se refiere al hecho de que las observaciones hechas en una determinada réplica no tienen una influencia sobre las observaciones hechas en otra réplica. Imaginemos por ejemplo que estamos estudiando la respuesta de los colibríes a la cantidad del néctar producido por las flores que éstos polinizan. Para medir esta respuesta establecemos dos parcelas adyacentes de 5 x 5 m. Una parcela es la control, mientras que la contigua es una parcela en donde se ha extraído todo el néctar de las flores. Tu variable respuesta es el número de visitas de los colibríes a las flores, y mides ésta en las dos parcelas. En la parcela control estimas un promedio de 10 visitas por hora, comparadas con sólo 5 visitas por hora en la parcela manipulada. Mientras que recoges los datos, observas que una vez que los colibríes llegan a la parcela manipulada, se marchan inmediatamente, y los mismos individuos visitan la parcela contigua (la control). Claramente, los dos grupos de observaciones no son independientes entre sí. Si la parcela control y la manipulada estuvieran más separadas entre sí en el espacio, las estimas del número de visitas promedio por hora podría haber sido distinta, y esto podría incluso modificar los resultados de nuestro estudio. Cuando las dos parcelas están próximas entre sí, la no independencia de las muestras aumenta la diferencia entre ellas, lo que podría provocar, entre otras cosas, un p -valor más bajo de forma espuria, y un error de tipo I (rechazo incorrecto de una hipótesis nula cierta). En otros casos, sin embargo, la no independencia podría disminuir las diferencias aparentes entre los tratamientos, contribuyendo al error de tipo II (aceptación incorrecta de una hipótesis nula falsa). Por desgracia, la no independencia de las muestras aumenta o disminuye tanto el p -valor como el poder estadístico de maneras no previsibles.

La mejor manera para salvaguardarnos de la no independencia de las réplicas es asegurarnos de que las réplicas dentro y entre tratamientos están separadas unas de otras en espacio y tiempo lo suficiente como asegurarnos de que unas no tengan un efecto sobre las otras. Sin embargo, en el campo de la ecología, como en muchos otros campos, casi todo tiene un efecto sobre casi todo a múltiples escalas espacio-temporales, y es por tanto muy difícil poder establecer una distancia o un espacio que garantice la no independencia de las

réplicas. Normalmente se recurre al sentido común y al conocimiento biológico que tengamos de nuestro caso de estudio. Pero ¿por qué no simplemente maximizar la distancia en tiempo o en espacio entre las muestras? Por varias razones. Primero, tomar muestras muy separadas unas de otras aumenta la necesidad de recursos económicos y humanos. Segundo, al separar mucho unas muestras de otras podemos incorporar nuevas fuentes de variación debido, por ejemplo, a la aparición de nuevos hábitats o ambientes. Queremos que nuestras réplicas estén lo suficientemente cerca como para tener un ambiente relativamente homogéneo pero lo suficientemente apartadas como para asegurar que la respuesta que estamos midiendo en cada una de ellas sea independiente del resto.

A pesar de la gran importancia que tiene este aspecto en el diseño experimental, muy raras veces es discutido en los artículos científicos. En la sección de Métodos de un trabajo científico es muy posible leer frases de este tipo: “*Medimos 100 plántulas seleccionadas de forma aleatoria en zonas soleadas. Cada plántula estaba separada a una distancia mínima de 50 cms de la plántula más próxima*”. Cuando lo que realmente ha guiado el diseño experimental es el hecho de: (1) querer disponer de 100 réplicas en un espacio limitado; y (2) no saber a ciencia cierta qué distancia de separación entre plántulas es la que garantizaría una total independencia en la variable respuesta. Los autores presuponen que como las plántulas de la especie estudiada tienen un tamaño muy pequeño, una distancia de 50 cms va a minimizar los efectos que tengan una sobre otra (ej. fenómenos de aleopatía, competencia por recursos, facilitación, etc.).

1.2. Factores de confusión

Cuando en el diseño experimental se superponen varios factores de forma simultánea, es muy difícil explicar los efectos por separado de cada uno de ellos sobre la variable respuesta. Volvamos al ejemplo de los colibríes. Tenemos ahora dos parcelas (una control y otra manipulada en dónde se ha extraído el néctar de las flores) separadas la una de la otra a una distancia suficientemente razonable como para asumir que no hay ningún tipo de dependencia entre ellas. Sin embargo, sin darnos cuenta hemos situado la parcela control en una ladera muy soleada y la parcela manipulada en un valle frío. El número de visitas promedio en la parcela manipulada sigue siendo menor que en la control, pero ahora ya no sabemos si ésto es debido al efecto del néctar o al hecho de que los colibríes disminuyen su tasa de actividad a bajas temperaturas. Por tanto no podemos discernir un efecto de otro sobre la variable respuesta. Podríamos incluso encontrar que un efecto anula al otro si, por ejemplo, hubiéramos situado la parcela control en el valle frío y la parcela experimental en la ladera soleada.

1.3. Replicación y aleatorización

La incorporación de replicación y aleatorización de las réplicas dentro de nuestro diseño experimental puede ayudar a controlar en gran medida los problemas de no independencia y la aparición de factores de confusión. Por

replicación, nos referimos aquí a la medición de la variable respuesta en varias muestras dentro de cada nivel del tratamiento o grupo de comparación. Por aleatorización, nos estamos refiriendo a la asignación aleatoria de los tratamientos o la selección de muestras.

Volviendo al ejemplo anterior, si seguimos los principios de replicación y aleatorización, estableceríamos varias parcelas control y varias parcelas manipuladas en vez de una sólo de cada (idealmente, un mínimo de 10 por cada tratamiento). La localización de estas parcelas en el área de estudio sería al azar, y la asignación del tratamiento (control o manipulación) a cada parcela también sería al azar.

¿Cómo disminuye la replicación y la aleatorización el problema de los factores de confusión? Tanto la ladera soleada como el valle frío, así como otras situaciones intermedias, tendrían por azar una o varias parcelas de cada uno de los tratamientos. Por tanto, el factor temperatura ya no se confundiría con el factor presencia de néctar en las flores, ya que todos los tratamientos de néctar ocurren bajo distintos niveles de temperatura. Como un beneficio adicional de este diseño experimental, podríamos comprobar de forma explícita el efecto de la temperatura (como una covariable en el modelo) sobre el número de visitas de los colibríes a las flores, e incluso la interacción entre la temperatura y la presencia de néctar. Por supuesto, si sabemos a priori que la temperatura tiene un efecto sobre la actividad de los colibríes, podríamos haber optado por realizar un diseño experimental distinto (ej. estratificado por zonas de temperatura) para comprobar específicamente esta hipótesis, pero desconociendo la influencia de éste u otros posibles factores de confusión, un diseño replicado y aleatorizado nos va a permitir minimizar los efectos confundidos de nuestro tratamiento con otras variables no medidas en nuestro estudio.

Es menos obvio como la replicación y la aleatorización reducen el problema de no independencia de las muestras. Después de todo, si las parcelas están por azar muy cerca unas de otras, las visitas de los colibríes a las flores de una parcela podrían no ser independientes de las visitas a las flores de otra parcela, por mucha replicación y aleatorización que hayamos hecho. Siempre que sea posible, hay que usar el sentido común y el conocimiento del caso de estudio para separar las réplicas por unas distancias mínimas de cara a evitar la no independencia de las muestras. Sin embargo, si no conocemos todas las posibles fuerzas que pueden causar dependencia, una disposición aleatoria de las muestras más allá de unas distancias mínimas pre-establecidas, asegurará que las distancias entre muestras sea variable. Algunas muestras estarán muy próximas entre sí y otras muy alejadas. El efecto de la dependencia será por tanto fuerte en algunos pares de muestras, más débil en otros, y no existirán en el resto. Tales efectos pueden cancelarse unos a otros y pueden reducir las probabilidades de que los resultados estén consistentemente sesgados por la no independencia.

2. Diseño factorial o ANOVA

Si tu variable predictora es categórica y tu variable respuesta es continua, tu diseño se denomina de tipo ANOVA (*Analysis Of Variance*). ANOVA también se refiere al análisis estadístico de este tipo de diseños de experimentos.

En terminología ANOVA, los **tratamientos** son las diferentes categorías de las variables predictoras que se utilizan en el estudio. En un estudio experimental, los tratamientos representan las diferentes manipulaciones que se llevan a cabo (ej. extracción del néctar de las flores y no extracción o control). En un estudio observacional, los tratamientos representan los diferentes grupos que se quieren comparar (ej. tipos de hábitat). Dentro de cada tratamiento, se realizan múltiples observaciones, y cada una de estas observaciones es una **réplica**.

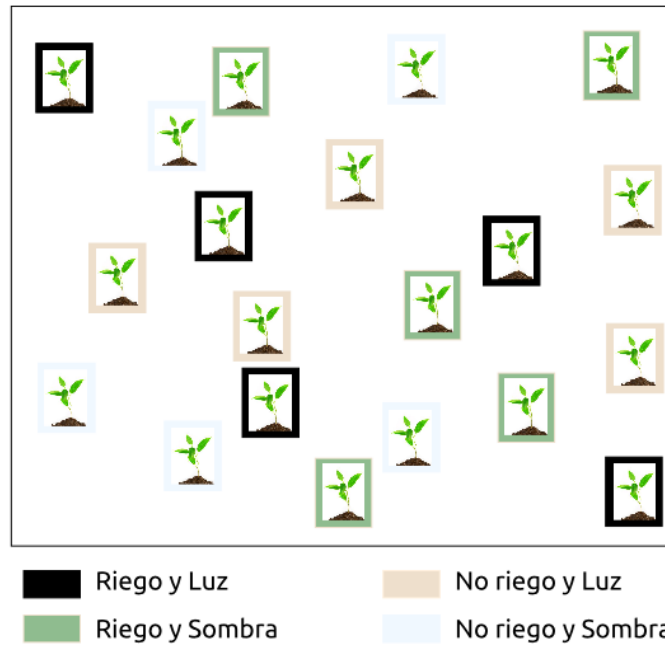
2.1. Supuestos de los diseños factoriales

Independencia En los diseños ANOVA estándar, cada réplica debe de ser independiente del resto de réplicas dentro del mismo tratamiento y entre tratamientos, esto es, no debe de tener un efecto claro sobre la respuesta observada en otras réplicas. Más adelante veremos algunos diseños ANOVA que permiten relajar este supuesto de independencia entre réplicas, como es el diseño por bloques o el diseño de medidas repetidas.

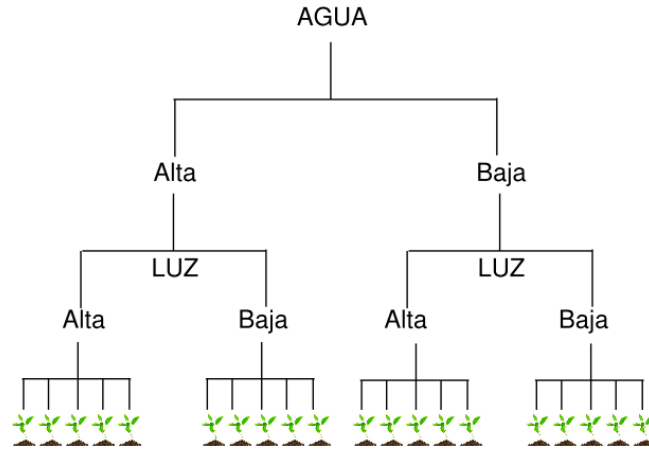
Cruzamiento Dentro del diseño ANOVA estándar, se puede distinguir entre diseños unifactoriales y diseños multifactoriales. Mientras que en los primeros los tratamientos representan variación en una única variable predictora o factor, en los diseños multifactoriales los tratamientos representan una combinación de los niveles de dos o más factores. El número de réplicas necesarias para llevar a cabo un análisis multifactorial crece muy rápidamente a medida que incluimos nuevos factores en el diseño. Si por ejemplo queremos ver el efecto de distintos niveles de nitrógeno (4 niveles) y fósforo (3 niveles) sobre el crecimiento de una determinada planta, necesitaríamos $4 \times 3 = 12$ tratamientos distintos. Si queremos disponer de un número de réplicas mínimo adecuado para comprobar nuestras hipótesis (10 réplicas por tratamiento) necesitaríamos al menos $12 \times 10 = 120$ réplicas. Si incluimos ahora un nuevo factor, por ejemplo, riego o no riego (2 niveles), requeriríamos al menos 240 réplicas para poder comprobar nuestras hipótesis. En los diseños ANOVA multifactoriales estándar, se asume que los factores son cruzados, es decir, que todos los niveles de un factor están representados dentro de todos los niveles del otro u otros factores. De esta manera es posible estimar si hay una interacción entre dos o más factores. Como veremos más adelante, algunos diseños ANOVA permiten incorporar factores que no son cruzados. A estos diseños se les denomina normalmente diseños anidados o jerarquizados.

Ejemplo 1 Se quiere ver cómo dos factores, luz y agua, afectan al crecimiento de una planta. Cada uno de estos factores tiene dos niveles (alto, bajo). Un diseño factorial completo con dos factores se diseñaría de tal manera

que tuviéramos varias muestras **independientes** (idealmente 10, aunque en la figura sólo se representan 5) en **todas las combinaciones posibles** de cada uno de los dos niveles de luz y agua (tratamientos), como muestra la siguiente figura. La aleatorización se realizaría espacialmente para la selección de las plantas o la ubicación de las parcelas de muestreo, pero también se realizaría en la asignación de tratamientos a cada una de estas plantas o parcelas.



Este diseño podría representarse de forma esquemática de la siguiente forma:



3. Otros diseños de experimentos

3.1. Diseño por bloques aleatorizados

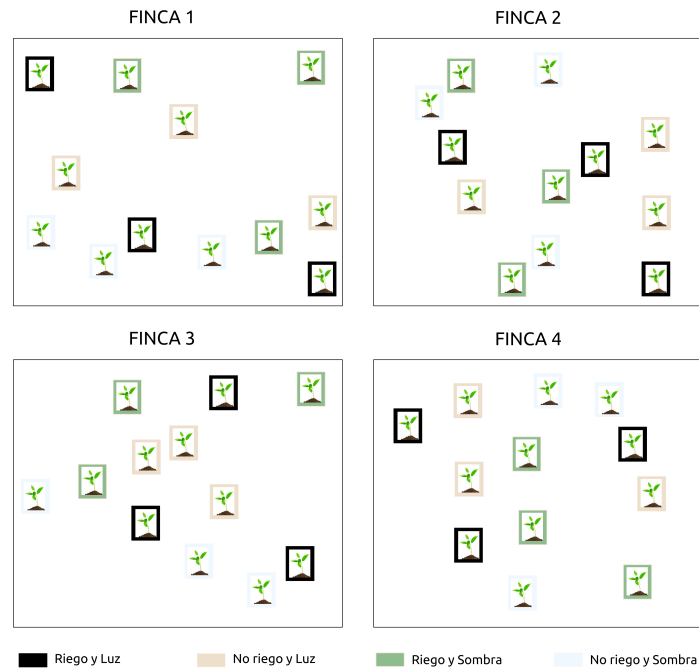
Una forma eficaz de incorporar la heterogeneidad ambiental en nuestro experimento es modificar el diseño factorial y utilizar un diseño por bloques aleatorizados. En un sentido estricto, el diseño por bloques aleatorizados es un diseño multifactorial, pero el segundo factor, al que se le denomina bloque, está incluido sólo para controlar la variación atribuible al diseño experimental (como consecuencia de la agregación de las muestras por bloques, violando por tanto el supuesto de independencia) y no es de interés para el investigador. Un bloque es un área o periodo de tiempo delimitado dentro del cual las condiciones ambientales son relativamente homogéneas. Los bloques pueden ser ubicados en el área de estudio de forma aleatoria o sistemática, pero deberían de disponerse de tal forma que las condiciones ambientales sean más similares dentro de los bloques que entre bloques.

Una vez que se han establecido los bloques, las réplicas deberán ser asignadas aleatoriamente a los tratamientos dentro de cada bloque, aunque en su forma más pura este tipo de diseños impone una restricción en la aleatorización: una única réplica de cada uno de los tratamientos debe ser asignada a cada bloque. Si todos los posibles tratamientos están representados dentro de cada bloque el diseño se denomina de **bloques aleatorizados completos**, pero si no todos los tratamientos están representados dentro de cada bloque el diseño se denominará de **bloques aleatorizados incompletos** (esto puede ocurrir con más frecuencia en estudios observacionales).

Los diseños por bloques aleatorizados pueden tener algunas desventajas. Una es que si los bloques son muy pequeños, se puede introducir no independencia de las réplicas por factores no controlados (ej. si el bloque está en una

pendiente las réplicas en la parte alta de la pendiente pueden tener una respuesta homogénea y diferencia con respecto a las réplicas situadas en la parte baja de la pendiente). Para solventar este problema se puede (y se debe) aleatorizar la ubicación de las réplicas dentro de cada bloque. Otra gran desventaja del diseño por bloques aleatorizados es que se asume que no hay interacción entre los bloques y los tratamientos. El diseño por bloques tiene en cuenta diferencias aditivas en la variable respuesta y asume que la relación entre las respuestas a cada tratamiento no varía entre un bloque y otro. Por ejemplo, si estamos investigando el efecto de tres fertilizantes (A, B y C) sobre el crecimiento de una planta, y replicamos el experimento en cinco bloques distintos, debemos esperar que si la mayor producción en el bloque I se observa con los fertilizantes $B > C > A$, esta misma relación se observará en el resto de bloques. Pero esto obviamente no tiene por qué ser así, y en ocasiones puede ser que esta relación cambie de unos bloques a otros (tendríamos que averiguar por qué se produce esta interacción entre el bloque y el factor). Por eso, algunos autores sugieren que el diseño tradicional por bloques aleatorizados no debe realizarse si no se lleva a cabo una replicación de los distintos tratamientos dentro de cada bloque. A veces esta replicación no es posible. En estas situaciones, el diseño por bloques aleatorizados simple va por lo menos a captar la componente aditiva de la variación ambiental.

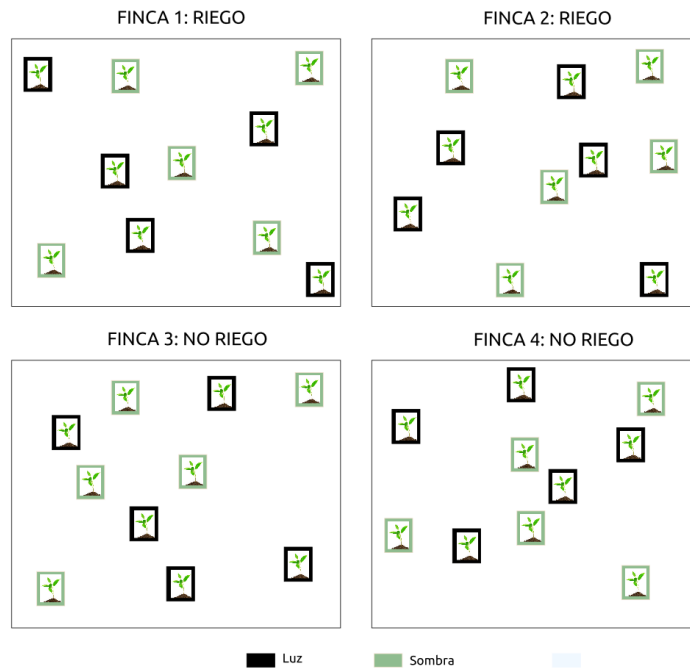
Ejemplo 2 Siguiendo con el ejemplo anterior, imaginemos ahora que las plantas que estamos cultivando en condiciones experimentales se encuentran repartidas en cuatro fincas. Dentro de cada finca hemos aleatorizado la ubicación de las réplicas y la asignación de los tratamientos a las réplicas. Podemos pensar que la situación geográfica de cada una de las parcelas (clima general, naturaleza del sustrato en dónde hacemos crecer las plantas, etc.) puede influir sobre la variable respuesta. Como esta variable no es de interés para el estudio pero queremos controlar como afecta a la variabilidad de la variable respuesta, la introducimos como un bloque en el análisis. Este sería un diseño por bloques aleatorizados en dónde existiría replicación de los tratamientos dentro de cada bloque, por lo que se podría además investigar si existe una interacción entre el bloque y los factores principales.



3.2. Split-plot

Los **split-plot** son una extensión de los diseños por bloques aleatorizados. Los diseños de tipo split-plot se originaron en el campo de la agronomía, pero hoy en día se encuentran con frecuencia en muchos otros ámbitos. Como su nombre indica, las parcelas experimentales son divididas en varias subparcelas. La parcela es sometida a los distintos tratamientos de uno de los factores. Hasta aquí, el diseño se corresponde perfectamente con el diseño por bloques aleatorizados. Lo que distingue a un diseño de tipo split-plot de uno por bloques aleatorizados es el hecho de que cada parcela o bloque es ahora sometido a los efectos de un segundo factor, permitiendo replicación de estos tratamientos entre las parcelas o bloques. Este tipo de diseños surgen como consecuencia generalmente de restricciones impuestas en las condiciones de experimentación, lo que impide una completa aleatorización del diseño.

Ejemplo 3 En el ejemplo anterior, imaginemos ahora que es muy caro llevar cisternas a cada una de las fincas para simular el riego. Por restricciones de experimentación, decidimos someter a riego a todas las réplicas dentro de dos de las cuatro fincas, y dejar sin riego a todas las réplicas de las otras dos. El factor luz sería aplicado de forma aleatoria dentro de cada finca (bloque). Todos los tratamientos del factor luz se aplicarían por tanto dentro de cada bloque, pero los tratamientos del otro factor (agua) serían aplicados a bloques completos.



Comparemos este diseño con el diseño de dos factores totalmente aleatorizado del Ejemplo 1. La principal diferencia es que, en el diseño de dos factores totalmente aleatorizados, cada réplica recibe las aplicaciones de los tratamientos resultantes de combinar los niveles de los dos factores de forma independiente. En el diseño de tipo split-plot, uno de los tratamientos se aplica a bloques enteros o parcelas, y el otro se aplica a las réplicas dentro de cada bloque.

Ejemplo 4 Queremos investigar el efecto de tres frecuencias de riego (diario, cada dos días, cada tres días) y cuatro variedades de semilla (A, B, C, D) sobre la producción de alfalfa. Bajo un diseño totalmente aleatorizado tendríamos que seleccionar parcelas de cultivo independientes entre sí y someterlas a todas las posibles combinaciones de los dos factores riego x variedad. Es decir que tendríamos que asignar parcelas a un mínimo de $3 \times 4 = 12$ grupos y, para tener réplicas dentro de cada grupo y poder ver el efecto de la interacción entre estos dos factores necesitaríamos al menos 24 parcelas de cultivo (aunque si nos atenemos al ideal de 10 réplicas por tratamiento, necesitaríamos al menos 120 réplicas). En experimentación agrícola esto supone una limitante muy importante. Una opción sería subdividir cada parcela de cultivo en doce subparcelas y aplicar en cada una de ellas una combinación de los niveles de los dos factores investigados. De esta manera con 6 parcelas tendríamos un total de $6 \times 12 = 72$ muestras, eso sí, no independientes entre sí. Una manera de analizar estos datos sería por medio de un anova factorial completo con un bloque (parcela, con 6 niveles).

Parcela 1				Parcela 2				Parcela 3			
R3 A	R2 B	R1 C	R3 D	R2 B	R3 C	R2 A	R3 D	R1 C	R2 B	R1 D	R3 A
R1 C	R3 A	R3 D	R1 B	R1 D	R1 A	R3 C	R2 B	R3 C	R3 A	R2 D	R2 B
R2 B	R1 D	R2 A	R2 C	R3 A	R2 D	R1 B	R1 C	R2 D	R1 A	R3 C	R1 B

Parcela 4				Parcela 5				Parcela 6			
R1 A	R3 D	R3 C	R3 B	R3 B	R1 A	R1 C	R2 D	R3 D	R1 B	R2 C	R3 A
R2 A	R2 C	R2 D	R1 B	R1 C	R3 A	R3 D	R3 B	R1 D	R2 A	R3 C	R1 B
R3 B	R1 D	R1 A	R2 C	R2 A	R2 C	R2 D	R1 B	R2 B	R3 D	R1 A	R2 C

Ahora imaginemos que no podemos implementar distintos tipos de riego dentro de cada parcela. Una opción sería subdividir cada parcela en cuatro subparcelas y aplicar en cada una de ellas una variedad de semilla. Y podríamos aplicar cada tipo de riego a dos parcelas distintas. Esto sería un diseño de tipo split-plot. Tendríamos ahora $6 \times 4 = 24$ muestras, no independientes entre sí, pero con un grado menor de independencia que en el caso anterior ya que hay un factor, que es el riego, cuyo efecto se mide entre parcelas.

Riego diario		Riego 2 días		Riego 3 días	
Parcela 1		Parcela 2		Parcela 3	
D	A	A	C	B	A
C	B	D	B	D	C
Parcela 4		Parcela 5		Parcela 6	
A	B	A	D	D	C
C	D	C	B	B	A

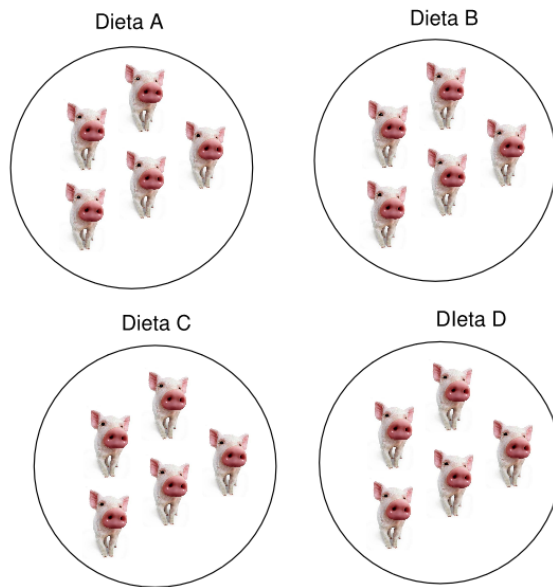
3.3. ANOVA de medidas repetidas

En todos los diseños que hemos visto hasta el momento, la variable respuesta es medida para cada réplica en un momento determinado al final del experimento. Un diseño de medidas repetidas se utiliza cuando se obtienen múltiples observaciones de una misma réplica en distintos momentos o periodos de tiempo. El diseño de medidas repetidas puede concebirse como un diseño por bloques aleatorizados, en donde las réplicas serían los bloques y a cada réplica se le aplicarían distintos tratamientos del factor principal en distintos momentos o periodos de tiempo.

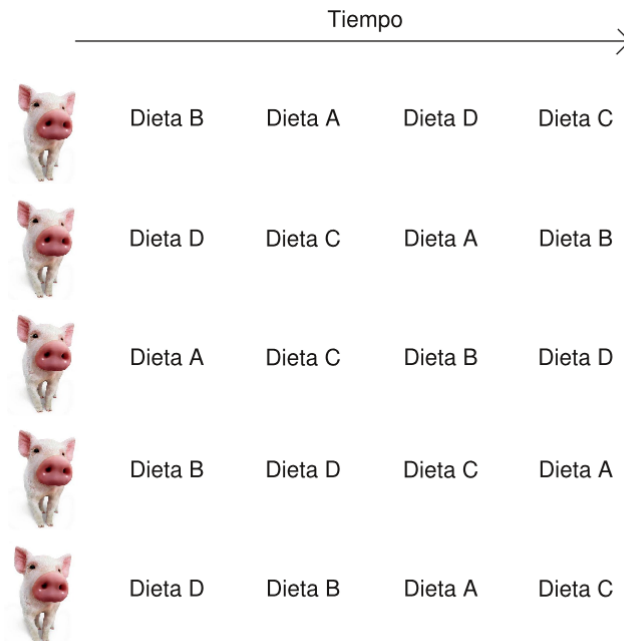
Los diseños de medidas repetidas se utilizan con mucha frecuencia en los campos de la medicina y la psicología, en donde observaciones repetidas son

tomadas en los mismos sujetos experimentales variando los tratamientos aplicados a los mismos. En la terminología de los diseños de medidas repetidas, los factores cuyos efectos se miden dentro de sujetos se denominan **factores intra-sujetos**, mientras que los factores cuyos efectos se miden entre sujetos se denominan **factores inter-sujetos**. En principio los diseños de medidas repetidas requieren de la existencia de, al menos, un factor intra-sujetos, que sería equivalente al bloque en el diseño por bloques aleatorizados. Si además de un factor intra-sujetos tenemos un factor inter-sujetos tendríamos un diseño equivalente al split-plot.

Ejemplo 5 Imaginemos una investigación diseñada para conocer el efecto de cuatro tipos de dietas sobre el engorde de los cerdos. Podemos optar por seleccionar tantos grupos de sujetos como dietas disponibles (cuatro) y someter a cada grupo a un único tipo de dieta. De esta manera tendremos un diseño con un factor (tipo de dieta, con cuatro niveles) y tantos grupos de sujetos como niveles del factor (cuatro). Para analizar los datos de este diseño podemos utilizar un ANOVA de un factor completamente aleatorizado.



En lugar de esto, podemos seleccionar un único grupo de sujetos y someterles de manera secuencial (dejando un intervalo de tiempo entre la aplicación de una dieta y otra) a los cuatro tipos de dietas distintas. En este caso seguiremos teniendo un diseño de un factor (tipo de dieta, con cuatro niveles), pero un sólo grupo de sujetos que hacemos pasar por las cuatro condiciones definidas por los niveles del factor (tendremos a todos los sujetos sometidos a todas las dietas). Es importante aleatorizar la aplicación de los tratamientos dentro de cada individuo (réplica) al igual que hacíamos con las réplicas dentro de cada bloque en el diseño por bloques aleatorizados. Para analizar los datos de este diseño podemos utilizar un ANOVA de medidas repetidas.



Ejemplo 6 Se quiere ver el efecto de 3 tratamientos de depuración de aguas residuales distintos sobre la producción de vertidos contaminantes al cauce de un río. Se prueban los tres tratamientos de manera consecutiva en 12 fábricas distintas y después de implementar cada uno de los tratamientos se mide la producción de vertidos contaminantes.

Ejemplo 7 Se está investigando el efecto de distintos tipos de sustrato (granito, cemento, pizarra) sobre el reclutamiento de percebes en granjas de producción acuícola. Un diseño de un factor totalmente aleatorizado implicaría seleccionar distintas réplicas sobre los tres tratamientos (sustratos). Como se trata de un experimento manipulativo, colocamos en distintos puntos de la costa 10 estructuras cuadradas de cada tipo de sustrato. Pasados varios meses medimos el reclutamiento de percebes. Imaginemos que ahora decidimos realizar mediciones cada 12 meses durante 3 años (3 mediciones). Podríamos incorporar el año como un factor intra-sujetos, por lo que tendríamos el equivalente a un split-plot pero en donde el bloque sería la réplica (estructura cuadrada muestreada) y éstas experimentarían los distintos niveles del factor tiempo (diseño de medidas repetidas)¹.

¹Si observáramos que el reclutamiento de percebes sigue una tendencia (positiva o negativa) lineal con el tiempo, también podríamos incluir el año como una variable continua (covariable) e incorporar un factor aleatorio (muestra) para especificar que la variabilidad se produce dentro de cada réplica.

Ventajas e inconvenientes del ANOVA de medidas repetidas Los diseños de medidas repetidas ofrecen varias ventajas. La primera es la eficiencia. Se requieren menos réplicas para comprobar los efectos del factor que un diseño totalmente aleatorizado. Además, los diseños de medidas repetidas permiten controlar la variabilidad debida a las diferencias que pueden existir de forma natural entre las réplicas o sujetos experimentales. Cuando las réplicas representan individuos (plantas, animales, o humanos), esto permite controlar de manera eficiente las diferencias entre individuos debidas al tamaño, edad, o a su historia de vida, todas las cuales pueden tener un efecto importante sobre la variable respuesta.

Los principales inconvenientes que podemos encontrar en un diseño de medidas repetidas son el **efecto de arrastre** (*carry-over effects*) y el **efecto de aprendizaje por la práctica**. El primero se produce cuando se administra un nuevo tratamiento antes de que haya terminado el efecto de otro tratamiento previamente administrado. El segundo ocurre cuando las respuestas de los sujetos a un tratamiento mejoran como consecuencia de los tratamientos previamente administrados bien sea como consecuencia del aprendizaje o como consecuencia de respuestas acumulativas. Por ello es conveniente aleatorizar los tratamientos dentro de cada sujeto, como en el Ejemplo 5.

El diseño de medidas repetidas (y también el diseño por bloques aleatorizados) hace un supuesto especial: el supuesto de **circularidad** o **esfericidad** para el factor intra-sujetos (en el caso de bloques aleatorizados, para el bloque). La circularidad en este contexto significa que la varianza de las diferencias entre cada dos tratamientos dentro de cada réplica es la misma. Si se produce cualquiera de los dos efectos descritos anteriormente o si existe alguna correlación temporal entre las observaciones hechas dentro de cada individuo, se incumplirá este supuesto, y se aumentará el error de Tipo I para los test F, por lo que habrá más probabilidades de rechazar incorrectamente la hipótesis nula cuando no deberíamos rechazarla. De hecho, en los diseños de medidas repetidas, se incumple este supuesto muy frecuentemente.

3.4. ANOVA anidado o jerarquizado

Hasta ahora todos los diseños de experimentos que hemos visto son **cruzados**. Los diseños cruzados son aquellos en donde existen todas las combinaciones posibles de los niveles de dos o más factores (ver Ejemplo 1). A veces puede ocurrir que tengamos algunos niveles de un factor combinados con unos niveles determinados de otro factor, sin que exista una representación de todas las posibles combinaciones de los niveles de los factores. Esto puede ocurrir cuando perdemos algunas réplicas del experimento, pero sobre todo ocurre cuando tenemos remuestreos dentro de cada réplica, algo que se conoce en la literatura científica con el nombre de **pseudorreplicación**.

3.4.1. Pseudorreplicación

La pseudorreplicación ocurre cuando analizas los datos como si tuvieras más grados de libertad de los que realmente tienes. Hay dos tipos de

pseudorreplicación: (1) **temporal**, que implica re-muestreos de los mismos individuos²; y (2) **espacial**, que implica que las muestras han sido tomadas en puntos próximos entre sí. La pseudorreplicación supone un problema grave porque uno de los supuestos más importantes de la mayoría de los análisis estadísticos es la independencia de los errores. La pseudorreplicación temporal tendrá errores no independientes porque las peculiaridades de los individuos remuestreados quedarán reflejadas en todas las medidas tomadas sobre ellos (es decir, las muestras procedentes de los mismos individuos estarán correlacionadas unas con otras). Las muestras tomadas en puntos próximos entre sí tendrán errores no independientes porque las peculiaridades de la localidad serán comunes a todas las muestras (por ejemplo, si varias muestras proceden de una zona con suelos muy fértiles, los valores de crecimiento en plantas serán todos muy altos y parecidos entre sí).

La pseudorreplicación es, por lo general, bastante fácil de distinguir. La pregunta que hay que hacerse es la siguiente: ¿Cuántos grados de libertad para los errores tiene el experimento realmente? Si un experimento de campo parece tener muchos grados de libertad, es posible que esté pseudorreplicado.

Tomemos un ejemplo procedente del control de plagas de insectos en plantas. Tenemos 20 parcelas, 10 fumigadas y 10 no fumigadas. Dentro de cada parcela hay 50 plantas. Cada planta es medida 5 veces para asegurar que la medición está bien hecha. Este experimento genera $20 \times 50 \times 5 = 5000$ valores. Hay dos tratamientos (fumigado, no fumigado), así que debe de haber un grado de libertad para el factor y 4998 grados de libertad para el término error.

Contemos ahora las réplicas (es decir, muestras independientes que experimentan el mismo nivel o combinaciones de los niveles de los factores) que hay en este experimento. Los remuestreos hechos sobre las mismas plantas (las cinco muestras por planta) no son réplicas. Las 50 plantas individuales medidas dentro de cada parcela tampoco son réplicas ya que es muy probable que las condiciones de cada parcela sean únicas y afecten por igual a cada grupo de 50 plantas muestreadas dentro de ellas independientemente de la aplicación o no de la fumigación. Hay 10 parcelas fumigadas y 10 no fumigadas y cada parcela va a proporcionar un único valor no independiente de la variable respuesta (por ejemplo, la proporción del área de las hojas comida por los insectos). Por tanto, habrá 9 grados de libertad dentro de cada nivel del factor, y $9 \times 2 = 18$ grados de libertad para el término error en el experimento en su conjunto. No es difícil encontrar ejemplos de pseudorreplicación parecidos a éste en la literatura científica. El problema es que, si no se identifica correctamente la existencia de pseudorreplicación, podemos acabar sacando conclusiones equívocas sobre la significación de los resultados (con 4998 grados de libertad para el término error, es casi imposible no detectar alguna diferencia estadísticamente significativa).

Hay varias cosas que se pueden hacer cuando tus datos están pseudorreplicados:

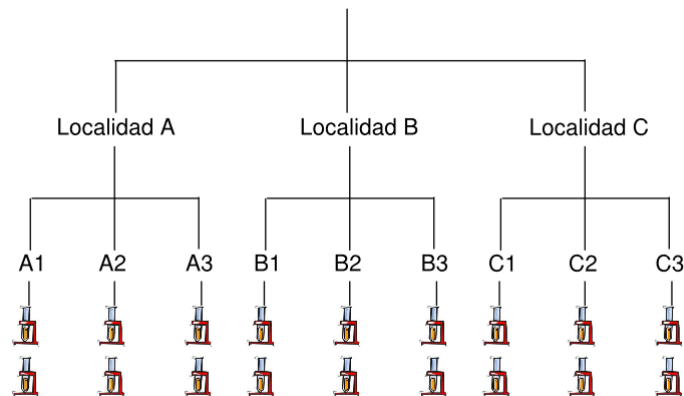
- Promediar la pseudorreplicación y llevar a cabo tu análisis estadístico con las medias;

²Cuando existe pseudorreplicación los individuos remuestreados representan un único nivel del factor. No confundir por tanto con los diseños de medidas repetidas, en donde cada individuo experimenta distintos niveles del factor.

- Hacer un análisis separado para cada periodo de tiempo en el caso de la pseudorreplicación temporal;
- Utilizar un análisis de series temporales o modelos mixtos.

Ejemplo 8 Se toman tres muestras de agua de tres pozos distintos. Con cada muestra se hacen dos determinaciones del contenido de fluoratos en el agua (miligramos/litro de agua). Se quiere ver si hay diferencias en los contenidos de fluorato entre los tres pozos.

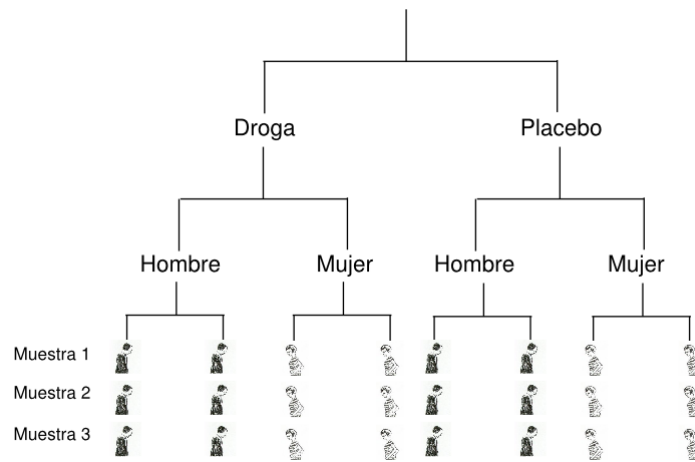
- Factor 1 = localidad (A, B, C).
- Factor 2 (anidado en F1) = muestra (1, ..., 9).



Ejemplo 9 Se quiere estudiar el efecto de una droga en hombres y mujeres. Se toman 8 individuos (4 hombres y 4 mujeres). A 2 hombres y a 2 mujeres se les suministra la droga, y al resto se les da un placebo. Se toma a cada individuo tres muestras de sangre y se mide la concentración de la droga en sangre.

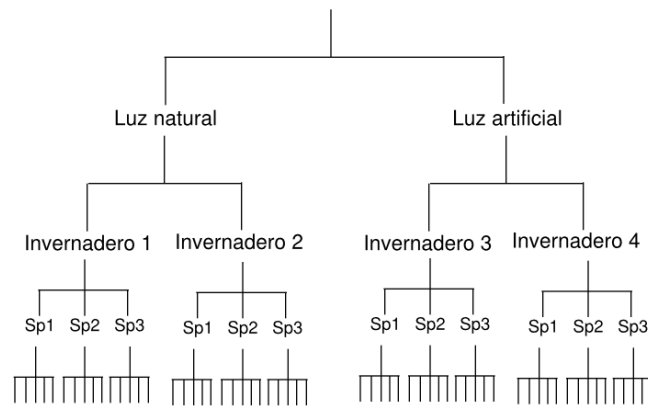
- Factor 1 = droga (control, droga)
- Factor 2 = sexo (hombre, mujer)

Estos dos factores son cruzados, porque tenemos una representación de todos los niveles del factor 2 (sexo) en todos los niveles del factor 1 (droga). Si sólo hubiéramos tomado una muestra de cada individuo, tendríamos muestras independientes entre sí y por tanto un diseño ANOVA de dos factores completamente aleatorizado. Pero... ¡tenemos tres muestras de cada individuo! Cada individuo debe ser tratado, por tanto, como un factor anidado dentro de cada una de las combinaciones de los niveles cruzados.



Ejemplo 10 Se quiere ver el efecto de la luz artificial sobre la producción de distintas especies de cereal. Se toman 4 invernaderos: 2 con luz natural, y 2 con luz artificial. En cada uno de los 4 invernaderos se plantan las mismas 3 especies de cereal. En cada invernadero se toman 5 muestras de cada especie y se calcula la productividad (kg/ha).

- Factor 1 = luz (2 niveles).
- Factor 2 (anidado en F1) = invernadero (cuatro niveles).
- Factor 3 = especie (tres niveles).



4. Modelos mixtos

Son una expansión de los modelos lineales generales que permite la inclusión de variabilidad correlacionada y no constante. Los modelos mixtos proporcionan, por tanto, la flexibilidad necesaria para modelar no sólo las

medias, sino también las varianzas y covarianzas de los datos. Los modelos mixtos no están relacionados con el diseño de experimentos pero sí nos va a permitir analizar diseños de experimentos que incumplan el supuesto de independencia, como los diseños por bloques aleatorizados, los diseños de tipo split-plot, los diseños de medidas repetidas o los diseños anidados. Los modelos mixtos son llamados así porque incluyen dos tipos de factores:

1. Factores **fijos**. Son todos aquellos factores cuyos niveles han sido definidos a priori. Las conclusiones que saquemos de nuestro estudio no pueden, por tanto, extrapolarse a otros niveles del factor que no hayan sido incluidos en el modelo. Afectan sólo a la media de y .
2. Factores **aleatorios**. Son todos aquellos factores cuyos niveles representan una muestra aleatoria de todos los posibles niveles del factor. Las conclusiones que saquemos de nuestro estudio se refieren a todos los posibles niveles del factor, aunque no estén recogidos de forma explícita en el experimento. Afectan sólo a la varianza de y .

A veces una variable explicativa representa una agrupación en el espacio o tiempo. Este tipo de diseños contradice los supuestos básicos de los modelos estadísticos estándar: la independencia de los errores. Los modelos mixtos tienen en cuenta esta no-independencia de los errores al modelar la estructura de la covarianza introducida por el agrupamiento de los datos. Una de las grandes ventajas de los modelos mixtos es que economizan grados de libertad típicamente utilizados por los niveles del factor. En vez de estimar una media para cada uno de los niveles del factor, el modelo de efectos aleatorios estima la distribución de las medias (normalmente como la desviación estándar de las diferencias de las medias de cada uno de los niveles del factor alrededor de la gran media). Los modelos mixtos son particularmente útiles en los casos en donde hay pseudorreplicación temporal (medidas repetidas o diseños anidados) o espacial (diseños anidados, por bloques aleatorizados o experimentos de tipo split-plot). En definitiva, los modelos mixtos nos van a permitir, por un lado, no desperdiciar grados de libertad en la estimación de los parámetros de cada uno de los niveles del factor de agrupación y , por otro, hacer uso de todas las mediciones que hemos realizado.

Un efecto aleatorio debería ser considerado como que proviene de una población de posibles efectos: la existencia de esta población de posibles efectos es un supuesto extra que debemos de considerar en el caso de los modelos mixtos. Así pues es más propio hablar de predicción de efectos aleatorios que de estimación de efectos aleatorios. Estimamos los efectos fijos a partir de los datos, pero hacemos predicciones sobre la población de la cual proceden nuestros efectos aleatorios. Los efectos fijos son constantes desconocidas que estimamos a partir de los datos. Los efectos aleatorios gobiernan la estructura de varianza-covarianza de la variable respuesta. Los efectos fijos son a menudo tratamientos experimentales que fueron aplicados bajo nuestra dirección, y los efectos aleatorios son variables que se distinguen por el hecho de que no nos interesan desde el punto de vista de los parámetros sino de la varianza que explican.

Otra diferencia importante entre efectos fijos y aleatorios es que los efectos fijos tienen niveles del factor **informativos**, mientras que los efectos aleatorios

tienen niveles del factor que **no** son **informativos**. La diferencia se ve mejor con un ejemplo. En los mamíferos la variable categórica *sexo* tiene dos niveles: macho y hembra. Para cualquier individuo que encuentres, el saber que es, por ejemplo, hembra, implica una gran cantidad de información sobre el individuo, y esta información se desprende de la experiencia recogida de muchos otros individuos hembra. Una hembra tendrá una serie de atributos (asociados con el hecho de ser hembra) sin importar la población a la que pertenezca el individuo en cuestión. Tomemos ahora una variable categórica como *genotipo*. Si tenemos dos genotipos en una población podríamos etiquetarlos A y B. Si tomamos dos genotipos de otra población diferente podríamos etiquetarlos igualmente A y B. En este caso, la etiqueta A no recoge ninguna información de referencia sobre el genotipo en cuestión, salvo que es probable que sea diferente al genotipo B. En el caso del *sexo*, el nivel del factor (macho o hembra) es informativo. *Sexo* es por tanto un factor fijo. En el caso del *genotipo*, el nivel del factor (A o B) no es informativo: *genotipo* es, por tanto, un factor aleatorio. Los efectos aleatorios tienen niveles del factor que son extraídos de una población mayor (potencialmente muy grande) en los cuales los individuos pueden diferir en su respuesta de muchas maneras, pero en dónde nosotros no sabemos exactamente cómo o por qué difieren.

4.1. ¿Efectos fijos o aleatorios?

A veces puede ser difícil decidir si una variable explicativa categórica tiene un efecto fijo (sobre la media de y) o un efecto aleatorio (sobre la varianza de y). A continuación se dan algunas claves que nos pueden ayudar a decidir esto:

- ¿Estoy interesado en el tamaño del efecto? Si la respuesta es **sí**, entonces **factor fijo**.
- ¿Es razonable asumir que los niveles del factor provienen de una población de niveles? Si la respuesta es **sí**, entonces **factor aleatorio**.
- ¿Hay suficientes niveles del factor a partir de los cuales estimar la varianza de los efectos de la población? Si la respuesta es **no**, entonces **factor fijo**. Zuur *et al.* (2007) estiman que al menos son necesarios 5 niveles del factor para poder considerar un factor como aleatorio, aunque es preferible que haya más de 10 niveles para esto. En cualquier caso si sólo hay 2 o 3 niveles siempre se ha de tratar un factor como fijo.
- ¿Son los niveles del factor informativos? Si la respuesta es **sí**, entonces factor fijo.
- ¿Son los niveles del factor simplemente etiquetas numéricas? Si la respuesta es **sí**, entonces generalmente factor aleatorio.

4.2. Modelos mixtos ¿una alternativa para representar y analizar otros diseños de experimentos?

Los modelos mixtos ofrecen una alternativa (pero no la única) para la representación y análisis de los diseños de experimentos más allá de los diseños

factoriales completamente aleatorizados, es decir, para todos los otros diseños vistos en este curso, incluyendo el diseño por bloques aleatorizados, los split-plot, los diseños de medidas repetidas y los diseños anidados o jerarquizados. Cualquier factor que represente una agrupación no independiente de las réplicas (como los diseños por bloques aleatorizados o los diseños de tipo split-plot) o cuya respuesta se mida dentro de los sujetos (como en el caso de los diseños de medidas repetidas o los diseños anidados) pueden ser analizados considerando la existencia de un factor de agrupación cuyos efectos son aleatorios.

Si el diseño es por bloques aleatorizados con un único factor el planteamiento para analizar los datos con modelos mixtos sería el siguiente:

$$y \sim X_1 + \text{random}(\text{Bloque})$$

Si el diseño es de tipo split-plot, con un factor aleatorizado entre bloques y otro aleatorizado dentro de bloques (Ejemplo 3), el planteamiento sería este otro:

$$y \sim X_1 * X_2 + \text{random}(\text{Bloque})$$

Si el diseño es de medidas repetidas con un único factor que opera intra-sujetos (Ejemplos 5 y 6) tendríamos un modelo teórico de este tipo:

$$y \sim X_1 + \text{random}(\text{Individuo})$$

Si además de tener un factor intra-sujetos tenemos un factor inter-sujetos (Ejemplo 7), el modelo sería este otro:

$$y \sim X_1 * X_2 + \text{random}(\text{Individuo})$$

Si tenemos un diseño anidado en dónde queremos probar el efecto de un factor sobre la variable respuesta, pero tomamos varias muestras de una misma réplica (Ejemplo 8), entonces tendríamos un modelo de este tipo:

$$y \sim X_1 + \text{random}(\text{Réplica})$$

Finalmente, si tenemos un diseño anidado con dos factores cruzados y varias muestras de las réplicas (Ejemplo 9), tendríamos lo siguiente:

$$y \sim X_1 * X_2 + \text{random}(\text{Réplica})$$

En dónde Bloque, Individuo o Réplica serían factores ya existentes o que tendríamos que crear con tantos niveles como bloques, individuos o réplicas haya. En la siguiente sesión se verá cómo analizar este tipo de diseños utilizando modelos mixtos en R.

5. Ejercicios

1. Se quiere ver cómo el efecto de la exposición (solana, umbría, rambla) y la localidad (Tárbena, Crevillente) afectan al tamaño de las hojas (en cm^2) de encina. Para ello se diseña un experimento en dónde se seleccionan aleatoriamente 24 individuos en tres zonas de solana, umbría y rambla respectivamente dentro de cada una de las dos localidades de estudio. A cada individuo se le miden 20 hojas para evitar datos que pudieran ser atípicos. ¿Cómo es la naturaleza de cada uno de los factores?
2. En un experimento agrícola se quiere investigar cómo la producción de un determinado cultivo depende de tres variables: irrigación (con dos niveles: regado o no); densidad de siembra (con tres niveles: baja, media y alta); y fertilización (con tres niveles: baja, media y alta). El diseño experimental se plantea de la siguiente manera: se eligen cuatro parcelas agrícolas independientes entre sí. Dichas parcelas son subdivididas en dos y a cada una de ellas se le aplica aleatoriamente uno de los dos tratamientos de irrigación. Cada una de estas subparcelas es a su vez dividida en tres partes y se aplica a cada una de ellas de manera aleatoria uno de los tres niveles de densidad de siembra. Por último, a cada una de estas subparcelas más pequeñas se las divide en tres y se aplica a cada una de ellas uno de los tres niveles de fertilización. ¿Cuántos factores hay y cuál es la naturaleza de cada uno de ellos?
3. En un experimento se quiere ver cómo el tipo de hábitat (bosque, matorral, pastizal) y distintos escenarios de cambio climático (control, aumento de la lluvia, disminución de la lluvia) pueden afectar diferentes respuestas en varias especies típicamente mediterráneas. Para ello se selecciona una zona de bosque, una de matorral y una de pastizal, todas próximas entre sí. En cada una de ellas se instalan 8 parcelas de 2×2 m. Cada parcela es subdividida en 9 subparcelas y se simulan en tres de cada 9 uno de los tres escenarios de cambio climático. Se plantan semillas de las plantas objeto de estudio y se miden las respuestas al cabo de un tiempo (porcentaje de germinación, crecimiento, producción de biomasa leñosa, etc). ¿Cuántos factores tenemos en este estudio y cómo son?
4. En un estudio se investiga la influencia de dos factores: herbivoría (con tres niveles: un control y dos niveles de intensidad) y lluvia (con tres niveles: un control y dos niveles de riego durante los meses de verano), sobre la cobertura de plantas herbáceas perennes en una dehesa mediterránea. Para ello se dispone de 9 parcelas de 2×2 m en dos bloques. Cada una de las parcelas de cada bloque es sometida a un nivel del tratamiento resultante de combinar los factores herbivoría y lluvia. El muestreo de cobertura se realiza en 6 cuadrats de 25×25 cm dentro de cada parcela. Además, se repite el muestreo durante 6 años consecutivos, para captar la variabilidad inter-anual propia de los sistemas mediterráneos, si bien se quiere captar la tendencia general y no la respuesta específica de las plantas en cada uno de los años del estudio. ¿Cuántos factores hay y cuál es la naturaleza de cada uno de ellos?

6. Referencias

- Gotelli, N.J., Ellison, A.M. (2004). A primer of ecological statistics. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA.
- Zuur, A.F., Ieno, E.N. & Smith, G.M. (2007). Analysing ecological data. Springer, New York.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009). Mixed effects models and extensions in ecology with R. Springer, New York.
- Crawley, M.J. (2007). The R Book. Wiley.