UNIVERSIDAD NACIONAL AGRARIA
Escuela de Post-Grado

Notas sobre el curso:

Estadistica Aplicada a la FORESTERIA II

Aplicación con el programa R

http://tarwi.lamolina.edu.pe/~fmendiburu/

2007

*** Preparado Por: Ing. Felipe de Mendiburu Delgado**

INDICE DE TEMAS

1. Principios basicos del diseño experimental
2. Tipos de experimentos
3. Relación de tipo de experimento con el diseño
4. Diseño Completo al Azar o de una Via
5. Cambio de escala
6. Diseño de Bloques Completos al Azar.
7. Diseño Cuadrado Latino: DCL
8. Analisis de Covariancia. DCA y Homogenidad de regresiones
9. Experimentos con factoriales cruzados y anidados.
10. Diseño en Parcela Dividida (Split plot)
11. Diseño con Bloques Divididos (Strip plot).
12. Experimentos factoriales sin repetición
13. Experimentos Repetidos en espacio y tiempo.
14. Experimento en campo de agricultores
15. Aplicaciones.
16. Referencia bibliográfica.

OBJETIVO

Presentar las tecnicas estadisticas necesarias para el diseño y analisis de experimentos eficientes para estudios de Investigación en las ciencias forestales; con enfasis en la interpretación y comprension de los resultados de los analisis estadisticos de datos experimentales.

Para el logro del curso, el Estudiante, al final del curso, debe ser capaz de:

- Diseñar, analizar e interpretar una variedad de experimentos.
- Mostrar competencia en el manejo de datos forestales.
- Usar paquetes estadisticos en computadoras.
- Presentar los resultados de investigación en la forma de una artículo científico o reporte.

RESUMEN

El presente documente describe el diseño experimental en su contexto general, destacando la importancia entre el diseño y el análisis estadístico. La participación del investigador en decidir que tipo de diseño debe utilizar en terminos de las unidades experimentales y los tratamientos y cuales son las fuentes de variación controladas. Se describe los diseños tradicionales como completo al azar bloques y cuadrado latino, así como otros diseños de mayor complejidad como experimentos factoriales anidados y cruzados, parcelas y bloques divididos, experimentos repetidos en espacio y tiempo, experimentos factorials sin repeticion y diseños en campo de agricultures en donde no es posible tener repetcion del tratamiento, adicionalmente se incluye el análisis de covariancia para la explicación de variables exogenas al experimento y que permiten disminuir el error. Se adiciona algunos casos de diseños aplicados y programas en R.

DISEÑO EXPERIMENTAL

El Diseño es una etapa fundamental de la experimentación, entiéndase por experimentación a toda investigación científica que se realiza por la repetición del mismo. El diseño comprende la forma de aplicar los tratamientos a las unidades experimentales y mediante un modelo estadístico cuantificar la variación debido a factores controlables. Entre los diseños más conocidos y utilizados en medios controlados es el completo al azar, en campo experimental los diseños de bloques, cuadrado latino, parcelas divididas, etc. En campo de agricultores puede planearse los diseños de bloques o diseños con balance incompleto, grupos balanceados, aumentados, etc. La incidencia de bosques en los microclimas, en el control de erosión, en la agricultura asociada, la adaptación de especies arbóreas, son temas que deben ser estudiados con la experimentación y determinar la incidencia en la producción, control de plagas, maleza, fertilizante, etc.

¿Porqué diseñar el experimento?

En primer lugar el investigador se formula una serie de preguntas, que espera tener respuesta al realizar el experimento, por ejemplo:

- ¿Cómo medirá el efecto de estudio?, ¿Cuáles serán las características a analizar?
- ¿Qué factores afectan a las características de estudio?.
- ¿Qué factores deben estudiarse?.
- ¿Cuántas veces debe realizar experimentos preliminares antes de conducir un experimento formal?.
- ¿Cuál seria el modelo de estudio para los datos del experimento?.

Muchas preguntas más se formularán y el investigador estará obligado a tener muchas consideraciones en el planeamiento del experimento.

¿Qué objetivo debe lograr el investigador al diseñar el experimento?.

- Conseguir toda la información relacionada al problema en estudio.
- Lograr un diseño simple y eficiente como sea posible.
- Optimizar los recursos de tiempo, dinero, personal y material experimental.

Por estas razones en necesario el asesoramiento de un especialista en estadística para las primeras etapas del proyecto de investigación.

1. PRINCIPIOS BASICOS DEL DISEÑO EXPERIMENTAL

El tema del diseño experimental ha recibido siempre mucha atención. Ahora el investigador dispone de nuevos métodos estadísticos para la solución de los problemas que involucra las condiciones del campo en área experimental y en campo de agricultores, laboratorio e invernadero. Se reconoce en general que, mediante el uso de diseños adecuados y análisis estadístico apropiado, las decisiones son confiables.

Gran parte del progreso reciente en la teoría y aplicación del diseño y análisis de datos se suman al esfuerzo para satisfacer las necesidades en la investigación agrícola o biológica y también otros campos de la investigación. Los principios básicos del diseño experimental, según se comprende, fueron desarrollados por R. A. Fisher y sus asociados en el Rothamsted Experimental Station en Inglaterra. Los tres principios básicos son: la randomización, la repetición y lo que denominó “control local”. Es de vital importancia para el investigador comprender la lógica de estos principios para diseñar experimentos eficaces (Figura 1).

Randomización

El principio de la randomización es único en la experimentación moderna. Según manifiesta Fisher, la randomización de los elementos a experimentar es esencial para la validez del error experimental y reducir al mínimo del sesgo en los resultados. También es una condición necesaria para el cumplimiento de supuesto respecto a las probabilidades asociadas con afirmaciones fiduciarias y pruebas de hipótesis (Figura 1).

Para justificar el arreglo aleatorio de las parcelas en los experimentos del campo, Fisher declara que los ensayos de uniformidad han establecido que la fertilidad del suelo no esta distribuido aleatoriamente, las parcelas vecinas tienden a ser parecidas que aquellas mas distantes. Es más, la distribución de fertilidad del Suelo rara vez o nunca esta sistemáticamente distribuida que podría representarse por una fórmula matemática.

Es importante recordar que el cálculo del error experimental depende de las diferencias de parcelas tratadas en forma similar. Tal cálculo será válido solo cuando pares de parcelas tratadas por igual no están tan cerca ni tan lejos que los pares de parcelas tratadas de otro modo. En general el principio de la randomización de las parcelas del campo es un requisito fundamental en el diseño de los experimentos.

Repetición

La heterogeneidad del suelo es la principal fuente de error en los experimentos de campo. Teóricamente, la heterogeneidad de suelo puede superarse hasta cierto punto mediante la repetición. Fisher ilustra las relaciones generales de repetición con el error experimental, según se muestra en la figura 1.

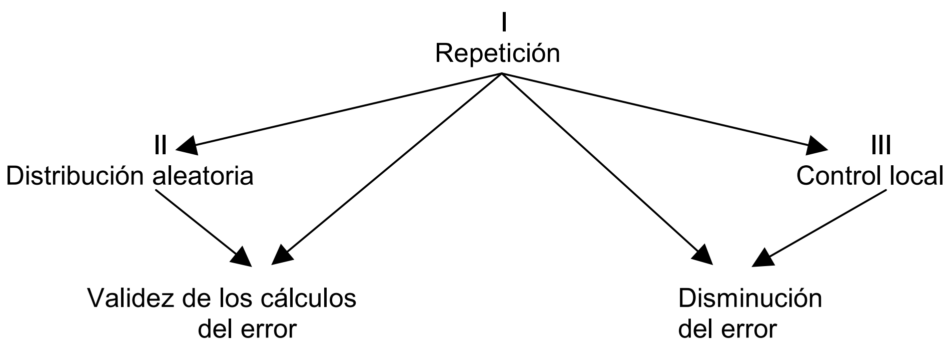


Figura 1. Diagrama de los principios en los cuales los métodos experimentales modernos han evolucionado.

Es evidente en este diagrama que la repetición funciona (a) a proveer un cálculo de la magnitud del error al cual se someten las comparaciones y (b) a disminuir el error experimental.

En lo que se refiere a la primera función de la repetición, las diferentes variedades o tratamientos en el experimento deben organizarse aleatoriamente para satisfacer la base matemática para un cálculo del error.

La segunda función de la repetición es la disminución del error, puede llevarse a cualquier grado de precisión, a condición de que un número suficiente de replicaciones se use junto con el control local. La variación de cualquier comparación de tratamientos disminuye directamente con un aumento del número de replicaciones, mientras la información sobre la comparación aumenta proporcionalmente. Es más, el error estándar medio es inversamente proporcional a la raíz cuadrada del número de replicaciones u observaciones, vale decir,

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Naturalmente el investigador quiere usar un número suficiente de replicaciones para lograr resultados confiables así como medir satisfactoriamente las diferencias. Es imposible determinar el número exacto de replicaciones necesarias para los experimentos. El número de replicaciones requeridas en los experimentos depende de tales factores como costo, trabajo, variabilidad del material, el tamaño probable de las diferencias de promedios y el nivel de significación deseada. En general, es imprudente usar menos de cuatro o cinco replicaciones en los experimentos del campo. Una regla útil es considerar por lo menos 10 grados de libertad en el error experimental. Sin embargo, debe reconocerse que hay límites prácticos más allá de los cuales el costo y la cantidad de trabajo incluido por un mayor número de replicaciones nos da una ganancia proporcional en la precisión.

Debe recalcarse a esta altura que la repetición y tamaño de parcela están estrechamente relacionadas, ya que una reducción del tamaño parcela permite un mayor número de parcelas en un área dada. El tamaño de parcela tambien esta relacionado al error experimental.

Control local

Un principio adicional del diseño experimental se llama el control local (figura 1). En el arreglo aleatorio de los tratamientos de un experimento, ciertas restricciones pueden invocarse para eliminar en parte la variación total que son irrelevantes al hacer las comparaciones.

El error experimental se controla más adecuadamente a través de la división del sitio para un experimento en campo, en varias áreas iguales denominados “bloques”. En los experimentos sencillos cada bloque o repetición contiene el mismo número de parcelas en donde las variedades o tratamientos a comparar se distribuyen de una manera aleatoria. Por lo tanto, el error experimental se reduce al mínimo debido al hecho de que la variación entre la parcela producido sobre el experimento es una parte cuantificable por las diferencias de los bloques.

Las diferencias entre las parcelas de un mismo tratamiento se deben en parte al error experimental y tambien al promedio de la diferencia entre las repeticiones. La variabilidad entre las repeticiones es irrelevante a la prueba experimental cuando cada tratamiento ocurre solo una vez en una repetición. Por consiguiente, la variación debida a las repeticiones se quita en general del error experimental. En consecuencia, la precisión del experimento es mayor cuando una cantidad grande de la variabilidad del total se quita de esta manera.

Hay muchos diseños de experimentos que incluyen arreglos diferentes de las parcelas, el más sencillo es el diseño de bloques y los arreglos cuadrados latinos aleatorizados.

2. TIPO DE EXPERIMENTOS

Los experimentos pueden clasificarse en varias tipos, basado en el número de factores o variables a estudiar en el momento. Un experimento formal es a veces precedido por una prueba preliminar.

Pruebas preliminares

Todas las pruebas preliminares son empíricas por naturaleza. Las pruebas brindan una oportunidad de detectar técnicas defectuosas, métodos inadecuados, etc. Una encuesta a veces se usa para una prueba preliminar. Información adquirida en tales pruebas forma una base para los diseños más eficaces de los experimentos formales. Por lo tanto, el experimento formal puede planificarse para eliminar muchas de las deficiencias observadas en la prueba preliminar para reducir el error experimental en los experimentos posteriores.

Experimentos con un factor

Estos experimentos son los más usado por los investigadores. Es sumamente recomendado debido a su sencillez. En estos experimentos otros factores se mantienen constantes o uniformes, en lo posible.

Estos experimentos son justificados cuando el tiempo, el material o equipo son limitados. Por ejemplo, un experimento puede ser determinar la mejor variedad de un cultivo, otro podría estar diseñado para determinar cual es el mejor fertilizante, mientras un tercer experimento separado para determinar las mejores prácticas culturales. Sin embargo, la información obtenida a partir de tales experimentos separados sería de utilidad limitada porque el investigador no podía determinar las interacciones posibles, es decir la interdependencia de los diferentes factores de variedad, fertilizante y prácticas culturales.

Experimentos factoriales

La experimentación factorial se debe principalmente a R. A. Fisher. En este tipo de diseño, dos o más factores pueden compararse en todas las combinaciones posibles con sus varios niveles. Por lo tanto, los resultados obtenidos no solo es la respuesta de los diferentes factores, sino también al mismo tiempo de las interacciones, es decir, la manera en que un cambio en un factor influye un cambio en otro. Es evidente que el experimento factorial combina dos o más experimentos sencillos de un factor en un solo experimento.

Hay dos ventajas del experimento factorial sobre los experimentos de un factor, vale decir, mayor eficiencia y mayor alcance, ventajas adicionales son las interacciones.

Ejemplos de los experimentos factoriales incluyen: (a) la relación de varios fertilizantes y métodos de la preparación de suelo, (b) la relación entre el fecha de plantar y la fecha de la madurez para producir, etc. Los experimentos factoriales constituyen un progreso muy importante en el desarrollo de los diseños de experimentos.

3. RELACIÓN DE TIPO DE EXPERIMENTO CON EL DISEÑO

Cada problema de investigación presenta limitaciones para considerar el diseño experimental apropiado. Sin embargo, una regla práctica es usar el diseño más sencillo que satisface los requisitos del experimento. El completo al azar, el bloque completo y en menor grado el cuadrado Latin son aplicados en campo, desde luego, esto no indica que los diseños más complejos se usan raramente.

Pruebas de rendimiento

Programas de mejoramiento de cultivos requieren la determinación de la capacidad de producción de las variedades superiores bajo diferentes suelos y las condiciones del clima. En consecuencia, la prueba de la variedad es probablemente el tipo más común de experimento del campo.

La densidad y fechas de siembra a veces se combinan con ensayos de la variedad, o pueden conducirse por separado. La prueba combinada permite un estudio de la respuesta de la variedad diferencial a las diferentes densidades y las fechas.

Para el fitopatólogo, los experimentos del campo incluyen pruebas de rendimiento de los tratamientos de semillas o los líquidos pulverizables y en polvo. Las pruebas de tratamiento de semillas quizá se combinen con densidad de siembra. El control de enfermedades experimenta frecuentemente las clases de productos químicos y sus dosis.

Al rociar y espolvorear en el material experimental, es importante considerar el efecto derivado de los productos químicos. Para reducir al mínimo este factor, quizá sea necesario insertar denominadas “filas del acarreo” entre las parcelas de tratamiento, esta practica incrementa el error experimental por efecto de la variabilidad del suelo y debe tomarse en cuenta. Este error puede compensarse por un mayor número de replicaciones.

En el trabajo de mejoramiento de plantas, con frecuencia es necesario probar un número grande de variedades o selecciones en un solo experimento. En tales condiciones el bloque completo aleatorizado no ejerce control suficiente sobre el error experimental. En estos casos, se recomienda el uso de diseños de grupos con bloques incompletos.

Experimentos culturales y de fertilizante

Los experimentos culturales incluyen tales variables como estaciones del año, los métodos de preparacion de siembra, la aplicación del agua de regadío, manejo del cultivo versus el uso de los herbicidas para el control de malezas, etc. Estos criterios a menudo se miden en el rendimiento. El bloque completo y los diseños cuadrados latinos aleatorizados en general son satisfactorios para las pruebas de esta clase. En aquellos ensayos que se usa maquinaria agricola o riego, es aconsejable usar algún tipo de diseño de parcela o bloque dividido.

La necesidad de probar fertilizantes deben ser en campo, por lo tanto, tales experimentos incluirían: (a) el efecto de ciertas formulaciones sobre el rendimiento, (b) las dosis de diferentes formulaciones en el rendimiento, (c) el efecto del tiempo de sembrado y los niveles de fertilización en el rendimiento y (d) el efecto de todas las combinaciones en el rendimiento.

Los diseños de bloques y cuadrados latinos aleatorizados son generalmente útiles
en los experimentos de fertilizante donde el número de factores interesados es razonablemente pequeño.

Para las pruebas de fertilizante donde 2 o más fertilizantes se aplican a 2 o más niveles, el experimento factorial es particularmente apropiado.

Experimentos en pastizales

El trabajo experimental en pastizal, el investigador puede desear: (a) determinar la cantidad de pasto producido en un área por mezclas diferentes de pastizal-gramíneas, (b) determinar la influencia de los fertilizantes en los pastizales en cuanto al rendimiento y la supervivencia de la especie de sabor agradable, (c) evaluar el efecto de algunos de tipos de manejo de pastizales en el rendimiento de diferentes mezclas de forraje.

El experimento factorial provee un diseño experimental apropiado a muchos tipos de experimentos en pastizales. Diseños alternados se han encontrado útiles para (a) el rendimiento los estudios sobre los pastizales anuales no resiembra, es decir, los cereales y ciertas leguminosas; y (b) los estudios de valor nutritivo para animales mayores.

Experimentos de rotación de cultivos

En rotación de cultivos u otros experimentos en los cuales se hace un estudio de los efectos residuales, es necesario cultivar todos los cultivos en la rotación de un año para obtener resultados fiables. Así, cada cosecha o fase de rotación se debe muestrear y ordenar los datos por cultivo y por fase de rotacion o años.

Experimentos con árboles

Los cultivos arbóreos son de larga vida y estan expuestos a mayor accidente que los cultivos anuales.

Debido a los requisitos extensos de espacio, el número de árboles por parcela es limitado, los resultados de las diferencias entre árboles individuales son con frecuencia muy variables, teniendo una fuente grande de error experimental.

Algunas respuestas pueden obtenerse en el mismo año al igual que los cultivos transitorios, pero en general se dispone de información acumulada para un analisis en el tiempo, esta es una caracteristica de los cultivos arbóreos.

Debido a las condiciones anteriores así como la posibilidad de tala o eliminación de árboles, el tipo de diseño experimental debe mantenerse sencillo, es decir tener un solo factor variable de estudio. La experiencia ha indicado que el diseño aleatorizado de bloques completos es sumamente útil con cultivos arbóreos

La experimentación con cultivos transitorios asociado al desarrollo de arboles o arbustos en areas tropicales pueden ser tratados como experimentos formales, asumiendo el tamaño de parcela lo correspondiente al cultivo, la distribucion de los arboles debe ser tal que el efecto sea estudiado en la forma que tenga beneficios tanto en la producción de los cultivos transitorios como la explotación de los arboles, por ejemplo: Efecto de sombra en la proteccion de la humedad del suelo para un cultivo; la produccion de semilla de arroz en rotación de stylosantes guianensis en asociación con árboles; efecto de especies arbustivas y leguminosas sobre la produccion anual de arroz; barbechos mejorados para la produccion de arroz, etc.

El uso de algunas leguminosas para mejorar la fertilidad del suelo, también es una práctica que puede experimentarse.

4. <u>Diseño Completo al Azar o de una Via</u>

Es el diseño más simple de aplicar, pero requiere ciertas condiciones respecto al material experimental. Es ampliamente utilizado en ambientes controlados como invernaderos, laboratorios, almacenes, etc.

Para aplicar este diseño se requiere que el material experimental sea homogéneo y suficiente para tener por lo menos 5 repeticiones por tratamiento. El tamaño de las unidades experimentales debe ser apropiado para el cultivo o tratamiento a aplicar.

El diseño consiste en aplicar los tratamientos aleatoriamente en todas las unidades experimentales con igual o diferente repetición, en lo posible tener igualdad de repeticiones.

Es recomendable tener pocos tratamientos y más repeticiones que muchos tratamientos y pocas repeticiones.

Por ejemplo, se evalúan niveles de fertilizante para producción de naranjas. Suponiendo que en un área extensa se disponen de 1000 árboles, se puede escoger un área de estudio homogénea, en la cual se identifica los árboles en un número suficiente, si son tres tipos de fertilizante, se puede tener por lo menos 5 árboles por tratamiento, distanciados para que un tratamiento no afecte a otro árbol bajo estudio.

El diseño mediante el programa R:

```
> library(agricolae)
> tratamientos <- c(“T1”,”T2”,”T3”)
> repeticiones <- c(5,5,5)
> design <- design.crd(tratamientos,repeticiones)
> design
>  design
      parcela trat
 [1,] "1"     "T3"
 [2,] "2"     "T2"
 [3,] "3"     "T2"
```

```
 [4,] "4"     "T2"
 [5,] "5"     "T1"
 [6,] "6"     "T3"
 [7,] "7"     "T1"
 [8,] "8"     "T2"
 [9,] "9"     "T1"
[10,] "10"    "T1"
[11,] "11"    "T2"
[12,] "12"    "T3"
[13,] "13"    "T3"
[14,] "14"    "T1"
[15,] "15"    "T3"
>
```

Bajo este esquema, se tendrá una sola fuente de variación controlada (fertilizante). Terminado el estudio, y escogiendo el fertilizante apropiado, puede ser recomendado para el uso en toda el área de los 1000 árboles.

Fuentes y grados de libertad para 3 tratamientos y 5 repeticiones.

Las fuentes generadas son los siguientes:

Fuentes	Gl.
Fertilizante	2
Error	12
Total	14

Las hipótesis de nulidad corresponde al efecto nulo del fertilizante en el rendimiento de los árboles.

Para las pruebas de comparación de medias para los fertilizantes puede utilizar pruebas de LSD, TUKEY, Contrastes o Dunnett.

Ejercicio. Leer el archivo “virus.txt” y realizar el ANVA hasta la comparacion multipe

```
# Analisis del diseño completo al azar
#
library(combinat)
datos <- read.table("virus.txt", header=TRUE)
#
attach(datos)
conjunto <-data.frame(virus=factor(virus),rdto=rdto)
attach(conjunto)
modelo <-aov(rdto ~ virus, conjunto)
summary(modelo)
gl <- gl.error(modelo)
cm <- cm.error(modelo)
compara<-LSD(rdto,virus,gl,cm)
```

```
Resultados:

            Df  Sum Sq Mean Sq F value    Pr(>F)
virus        3 1170.21  390.07  17.345 0.0007334 ***
Residuals    8  179.91   22.49
```

```
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Prueba LSD

Alpha = 0.05

Relacion de tratamientos
-------------------------
1  :  cc
2  :  fc
3  :  ff
4  :  oo

Promedios
-------------------------
1  :  24.4
2  :  12.86667
3  :  36.33333
4  :  36.9

Comparacion entre tratamientos
Trat.  Sign   Diff    LSD
-------------------------
1 - 2  *     11.53  8.929
1 - 3  *     11.93  8.929
1 - 4  *      12.5  8.929
2 - 3  *     23.47  8.929
2 - 4  *     24.03  8.929
3 - 4  ns   0.5667  8.929
```

5. <u>Cambio de escala.</u>

Para el analisis de variancia se necesita que los datos esten normalmente distribuidos, sin embargo en aquellos casos que la respuesta no es normal, por ejmplo valores discretos (contadas), proporciones o porcentajes respecto de un total; es necesario aplicar al valor original, un cambio de escala mediante las siguientes transformaciones:

Logaritmica. (base 10 ó e).- Cuando los datos presentan variancias heterogeneas entre grupos de tratamientos. Las variancias son proporcionales a sus promedios.

L = log(Y)

Cuando los datos puedan presentar efectos multiplicativos.

L = Log(Y+k), K=1 o un numero entero mayor.

Raiz Cuadrada.- Cuando los datos son discretos, por ejemplo numero de manchas en una hoja,
Número de bacterias en una placa.

$$R = \sqrt{Y + 0.5}$$

$R = \sqrt{Y}$

Angular.- Cuando la variable se expresa en porcentajes y esta dispersa en una escala de 0 a 100.

$A = Arc\sin\sqrt{Y/100}$

Si la funcion esta en radianes, se debe multiplicar el resultado por 180/3.1415 para tener la respuesta en angulos en la escala de 0 a 90

Cuando los % son menores del 20% o mayores del 80% en todos estos casos, solo se debe aplicar la raiz cuadrada. En el caso que solo este en el intervalo entre 20 a 80 tambien puede aplica la raiz cuadrada.

En todos los casos el Coefiente de variación se ve disminuida, esto no significa que para bajar el CV. se debe aplicar un cambio de escala. La transformación se debe aplicar una sola vez. Si tiene problemas con los analisis, se sugiere probar primero la normalidad de los errores, luego la no aditividad de los efectos del modelo. Finalmente utilice pruebas noparametricas para evaluar sus tratamientos de no cumplir los supuesto importantes del Analisis de Variancia.

6. Diseño de Bloques Completos al Azar.

Es similar al diseño de doble vía, un factor es aleatorio y corresponde a los bloques.

Para este diseño, el control local consiste en formar grupos de unidades homogéneas llamadas bloques, los tratamientos corresponden al otro factor que se supone fijo en algunos casos podría ser aleatorio, en cualquier caso, estos son asignados al azar en cada bloque. Las repeticiones de los tratamientos son los bloques.

Ejemplo: Estudio de tres especies maderables y una leguminosa como sombra para un cacaotal establecido, así como su efecto sobre la producción de cacao en 5 repeticiones. Se mide el DAP y el rendimiento de Cacao.

Factor de estudio: 4 tratamientos (3 especies maderables y una leguminosa)

El diseño puede realizarlo con el programa R:

```
> tratamientos <- c("LAUREL","GUABA","ROBLE","TERMINALIA")
> repeticiones <- 5
> design <- design.dbca(tratamientos,repeticiones)
> design

      parcela bloque trat
 [1,] "1"     "1"    "TERMINALIA"
 [2,] "2"     "1"    "LAUREL"
 [3,] "3"     "1"    "GUABA"
 [4,] "4"     "1"    "ROBLE"
 [5,] "5"     "2"    "GUABA"
 [6,] "6"     "2"    "TERMINALIA"
 [7,] "7"     "2"    "ROBLE"
 [8,] "8"     "2"    "LAUREL"
 [9,] "9"     "3"    "LAUREL"
[10,] "10"    "3"    "ROBLE"
[11,] "11"    "3"    "TERMINALIA"
[12,] "12"    "3"    "GUABA"
```

```
[13,] "13"    "4"    "LAUREL"
[14,] "14"    "4"    "TERMINALIA"
[15,] "15"    "4"    "ROBLE"
[16,] "16"    "4"    "GUABA"
[17,] "17"    "5"    "GUABA"
[18,] "18"    "5"    "LAUREL"
[19,] "19"    "5"    "ROBLE"
[20,] "20"    "5"    "TERMINALIA"
```

Para el caso planteado, las fuentes de variación controlada son dos: Variación debido al efecto de los tratamientos y variación debido al bloqueo, esta ultima variación corresponde a las diferencias entre las repeticiones que permiten disminuir el error experimental.

Fuentes	Gl.
Bloques	4
Especies	3
Error	12
Total	19

Las hipótesis de nulidad corresponde al efecto nulo de las especies sobre la producción de cacao.

Para las pruebas de diferencias entre especies, se puede utilizar pruebas de LSD, TUKEY, Contrastes o Dunnett.

Estudio de dos factores aleatorios.

Por lo general no se realiza un diseño experimental para estos estudios, la información es recolectada mediante muestras de las unidades afectadas y se estudia en un modelo lineal mediante el análisis de variancia.

Por ejemplo se dispone de muchos clones (100 o más) de un cultivo y se quiere evaluar si presentan diferencias estadísticas en el rendimiento y para ello se utiliza un diseño de bloques al azar, 5 bloques. Como son muchos los tratamientos, se selecciona 8 clones al azar del conjunto de clones y se aplica en el campo, de acuerdo a las exigencias del diseño de bloques.

Para el caso planteado, las fuentes de variación controlada son dos: Variación debido al efecto de los clones y variación debido al bloqueo.

Fuentes	Gl.
Bloques	4
Clones	7
Error	28
Total	39

Las hipótesis de nulidad corresponde a la variancia nula del efecto del clon.
Los factores aleatorios considerados en investigación corresponden a las repeticiones, años, días, meses, sitios y selección de tratamientos al azar. En todos estos casos no se estudia las comparaciones de promedios, solo interesa mediar el efecto en la variación de los datos.

7. <u>Diseño Cuadrado Latino: DCL</u>

En general, este diseño se aplica en el campo, donde el factor suelo obliga a formar grupos homogéneos en dos direcciones, puede ser pendiente u otras característica que pueda afectar al material experimental.

El diseño consiste en agrupar las unidades experimentales en dos direcciones (filas y columnas) y asignar los tratamientos al azar en las unidades, en una forma que en cada fila y en cada columna estén todos los tratamientos.

El total de unidades experimentales es igual al cuadrado del numero de tratamientos. Para 4 tratamientos se requiere 16 unidades.

Este diseño tiene algunas limitaciones en cuanto a los grados de libertad del error y para tener una mejor estimación del error, es recomendable doblar los tratamientos, es decir usar los mismos tratamientos como si fuesen otros y en el análisis estadístico hacer las debidas comparaciones que correspondan. En otros casos, puede ser factible hacer mas cuadros latinos y hacer un análisis combinado de experimentos repetidos.

El nombre de cuadrado Latino se debe a R.A. Fisher [The Arrangement of Field Experiments, J. Ministry Agric., 33: 503-513 (1926)]. Las primeras aplicaciones fueron en el campo agronómico,.

Formación de cuadrados latinos.
Suponga 4 tratamientos A, B, C y D, con estos tratamientos se pueden formar 4 cuadros diferentes llamadas típicas o estándar (en la primera fila y en la primera columna se tiene la misma distribución.

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

De cada cuadro se obtienen 144 formas diferentes, en total son 576 cuadros diferentes.

La siguiente tabla permite relacionar él numero de cuadros en función del tamaño.

Tamaño del cuadrado	Numero de formas Típica	Valor de n!(n-1)!	Número total de cuadrados diferentes
3 x 3	1	12	12
4 x 4	4	144	576
5 x 5	56	2880	161280
6 x 6	9408	86400	812851200

n = tamaño del cuadro.

El análisis y el diseño se ajustan a la siguiente distribución de los grados de libertad.

Fuentes	Gl	Ejemplo
Filas	n-1	3
Columnas	n-1	3
Tratamientos	n-1	3
Error	(n-1)(n-2)	6
Total	$n^2 - 1$	15

La nulidad del efecto del tratamiento se prueba con la relación: CM(tratamientos) dividido por el CM (error experimental), si este valor es mayor o igual a un valor de la distribución de F, (Probabilidad $\geq$ 0.05) con grados de libertad de n-1 y (n-1)(n-2), se rechaza la nulidad y se acepta que el efecto de tratamientos es significativo.

En la practica, con el uso de programas en computadora, es posible hallar la probabilidad de rechazo de la hipótesis de nulidad (Pr >F), llamado riesgo, si este valor es más pequeño que 0.05, los tratamientos muestran diferencia significativa en al menos un par de tratamientos.

Diseño DCL mediante el programa R.

Considere los siguientes Híbridos de Melon como tratamientos para un estudio.

V1 : Mission
V2 : Mark
V3 : Topfligth
V4 : Hales Best Jumbo

```
> tratamientos<- c("Mission","Mark","Topfligth","Hales Best Jumbo")
> design <- design.dcl(tratamientos)
> design

      parcela fila columna trat
 [1,] "1"     "1"  "1"     "Hales Best Jumbo"
 [2,] "2"     "1"  "2"     "Mark"
 [3,] "3"     "1"  "3"     "Topfligth"
 [4,] "4"     "1"  "4"     "Mission"
 [5,] "5"     "2"  "1"     "Mark"
 [6,] "6"     "2"  "2"     "Topfligth"
 [7,] "7"     "2"  "3"     "Mission"
 [8,] "8"     "2"  "4"     "Hales Best Jumbo"
 [9,] "9"     "3"  "1"     "Topfligth"
[10,] "10"    "3"  "2"     "Mission"
[11,] "11"    "3"  "3"     "Hales Best Jumbo"
[12,] "12"    "3"  "4"     "Mark"
[13,] "13"    "4"  "1"     "Mission"
[14,] "14"    "4"  "2"     "Hales Best Jumbo"
[15,] "15"    "4"  "3"     "Mark"
[16,] "16"    "4"  "4"     "Topfligth"
```

8. Analisis de Covariancia

Es una metodologia estadistica que utiliza el análisis de regresión y el análisis de variancia para tener un mejor control del error experimental. Esto es posible solo si se dispone de una o mas variables externas al experimento y se tiene informacion para cada unidad experimental del valor de estas variables.

El análisis de covariancia se puede aplicar a cualquier diseño experimental. Mediante el ANVA se descompone la variación total de la variable "Y" en factores controlables y no controlables, en el factor no controlable se considera el error experimental. En el error experimental están todos aquellos factores y variables que no pudieron ser medidos o simplemente no se midieron. Sin embargo, si una variable es factible de medir en cada unidad experimental, ésta debe ser considerada en el modelo, a menos que se pruebe estadísticamente que no tiene ningún efecto. El Análisis de Covariancia (ANCOVA) permite el estudio de estas variables externas (concomitantes), si deben o no ser consideradas en el modelo y en que forma se las controla.

Dentro de los posibles usos del ANCOVA están:

i. Control de variables externas que implica una disminución del error que se traduce en una mayor precisión del análisis.
ii. Ajuste de las medias de tratamientos de la variable dependiente (Y) por las diferentes variables independientes (concomitantes).
iii. Ayudar en la interpretación de los datos, especificamente en la naturaleza del efecto de los tratamientos.

Algunos ejemplos de aplicación :

- Altura inicial de plantulas (X) y el crecimiento en vivero al termino de un periodo, la respuesta Altura final (Y)
- El peso inicial (X) de animales relaciona al peso final (Y), cuando estos animales están sujetos a diferentes raciones. Se estudia el efecto de las raciones a través de los pesos observados.
- EL número de plantas (X) por parcela. Se estudia el rendimiento total (Y) de la parcela.
- El Rendimiento (X) de las parcelas en una producción anterior y el rendimiento (Y) de las mismas parcelas al finalizar el experimento. El estudio consiste en comparan variedades de un determinado cultivo.
- La incidencia de plagas (X) en el rendimiento de algunas variedades, el estudio es comparar las variedades.

En cada caso, se entiende que la variable X tiene un efecto en la variable Y, sin embargo esta dependencia deberá ser probada estadísticamente mediante el ANALISIS DE REGRESION.

Supuestos para el ANCOVA

Los supuestos requeridos para que sea válido el ANCOVA son:

i. Independencia de los errores.
ii. Normalidad en las variables aleatorias.
iii. Variancias homogéneas.
iv. Aditividad de los efectos involucrados en el modelo.
v. La variable X fija.
vi. La variable X no esta influenciada por los tratamientos.
vii. La variable Y está relacionada con X en forma lineal.

Los modelos estadísticos para los siguientes diseños son expresados como:

Diseño completo al azar.

$$Y_{ij} = \mu + \tau_i + ß(X_{ij} - \bar{X}) + \varepsilon_{ij} \qquad \begin{array}{l} i=1,2,...,t \\ j=1,2,...,r_i \end{array}$$

Diseño bloques completo al azar.

$$Y_{ij} = \mu + \tau_i + B_j + ß(X_{ij} - \bar{X}) + \varepsilon_{ij} \qquad \begin{array}{l} i=1,2,...,t \\ j=1,2,...,r \end{array}$$

En todos los modelos, ß es un parámetro que representa el coeficiente de regresión. Se supone que ß es distinto de cero, lo cual debe probarse, caso contrario en el modelo se elimina el termino afectado por ß.

ANCOVA en DCA.

En el primer modelo planteado, cada observación del experimento expresa el efecto del tratamiento y el efecto de la variable independiente (X), en forma lineal. El efecto de la variable independiente esta definido por:

$$ß(X_{ij} - \bar{X})$$

Para el análisis del experimento bajo el modelo DCA, deben estimarse los parámetros, luego realizar las pruebas de hipótesis:

- Prueba del efecto de regresión de X en Y, en la que no intervienen el efecto de tratamientos.
- Prueba de los tratamientos involucrados.

Las hipótesis a formularse en ANCOVA son:

Para la regresión:

H_p: ß = 0
H_a: ß $\neq$ 0

Para los tratamientos:

H_p: $\tau_i = 0$ para i=1,2,....,t
H_a: $\tau_i \neq 0$

Para realizar ambas pruebas de hipótesis se deben calcular las sumas de cuadrados y sumas de productos de las variables X e Y, mediante las fórmulas:

SUMAS DE CUADRADOS Y PRODUCTOS

Variable X: S_{xx} = SCTotal(X) = $\Sigma\Sigma X^2_{ij} - X^2../r.$
T_{xx} = SCTrat (X) = $\Sigma X^2_{i.}/r_i - X^2../r.$
E_{xx} = SCError(X) = $S_{xx} - T_{xx}$

Variable Y: S_{yy} = SCTotal(Y) = $\Sigma\Sigma Y^2_{ij} - Y^2../r.$
T_{yy} = SCTrat (Y) = $\Sigma Y^2_{i.}/r_i - Y^2../r.$
E_{yy} = SCError(Y) = $S_{yy} - T_{yy}$

Producto XY: S_{xy} = SPTotal(XY) = $\Sigma\Sigma X_{ij}Y_{ij} - X..Y../r.$
T_{xy} = SPTrat (XY) = $\Sigma X_{i.}Y_{i.}/r_i - X..Y../r.$
E_{xy} = SPError(XY) = $S_{xy} - T_{xy}$

Las sumas de productos pueden ser valores negativos.

PRUEBA DE HIPOTESIS H_p: ß=0 vs H_a: ß $\neq$ 0

b = est(ß) = E_{xy}/E_{xx}
SC regresion = b E_{xy}
SC residual = E_{yy} - SCregresión

Mediante la prueba de F se puede probar la hipotesis planteada

Fc = CM(regresion)/CM(reridual).

Si la hipótesis nula es cierta, F_c se distribuye como una F con (1, r.-t-1) grados de libertad, lo que implica que ß=0, es decir X no tiene relacion con Y, por lo tanto X no será considerado para el análisis de tratamientos.

Si Fc es mayor o igual a $F\alpha(1,n.-t-1)$ se rechaza la H_0: ß=0. Se afirma estadísticamente que ß es distinto de cero, implica que X tiene efecto sobre la respuesta de Y, y será considerado en el análisis de la variancia para la prueba de tratamientos.

PRUEBA DE HIPOTESIS H_p: τ_i=0 vs H_a: $\tau_i \neq 0$

Si ß es distinto de cero, se debe continuar con el siguiente análisis de variancia:

CUADRO DE ANCOVA

Fuentes	Gl	ΣX²	ΣXY	ΣY²	ΣY² -(ΣXY)²/ ΣX²	Gl	CM	
Tratamientos	t-1	Txx	Txy	Tyy				
Error	r.-t	Exx	Exy	Eyy	Eyy-(Exy)²/Exx	r.-t-1	CM Residual	
Trat+error	r.-1	Sxx	Sxy	Syy	Syy-(Sxy)²/Sxx	r.-2		
Tratamiento Ajustado	t-1				Diferencia (trat+error) y (error)	t-1	CM(Trat. ajustado)/(t-1)	Fc

El Valor de Fc es obtenido por el cociente entre el CM tratamiento Ajustado y CM residual (error ajustado).

Si la H_p es cierta, F_c sigue una distribución de F con grados de libertad (t-1) y (r.-t-1) para el numerador y denominador, respectivamente. La prueba de F permite probar la significación de los tratamientos mediante la relación:

Fc = CM trat ajustado / CM error ajustado.

CM error ajustado = CM residual

Si la hipótesis nula H_p: τ_i=0 es cierta, el efecto de tratamientos es no significativo.

Si Fc · $F\propto$(t-1,r.-t-1) se rechaza la H_p: τ_i=0. Se afirma que los tratamientos son significativamente diferentes al nivel $\propto$.

COMPARACION DE PROMEDIOS AJUSTADOS.

Si los tratamientos resultan significativamente diferentes, la prueba de DLS puede ser usada para comparar los promedios. La desviacion estándar de diferencia de promedios ajustados es:

$$Sd = \sqrt{CMe\left[\frac{1}{r_i} + \frac{1}{r_j} + \frac{\left(\overline{X}_i - \overline{X}_j\right)^2}{Exx}\right]}$$

y los promedios ajustados.

$\overline{Y}_i.(\text{ajustado}) = \overline{Y}_i. - b(\overline{X}_i. - \overline{X}..)$ para i=1,2,...,t

b = coeficiente de regresión.

La diferencia de dos promedios 1 vs 2 es:

Diferencia = $|\overline{Y}_1.(\text{ajust})-\overline{Y}_2.(\text{ajust})|=|\overline{Y}_2. -\overline{Y}_2. -b(\overline{X}_1.-\overline{X}_2.)|$

El valor DLS = $T\alpha(r.-t-1)\ S_d$

Los grados de libertad corresponden al error experimental ajustado.

El procedimiento en R, mediante la funcion lm: lineal model:

se considera como variables “x” e “y” y como tratamientos un factor.

lectura de los datos, almacenados en “conjunto”
> modelo <- lm(y ~ tratamientos + x , conjunto)
> anova(modelo)
> summary(modelo)

Este procedimiento realiza lo siguiente:

- Regresion
- ANCOVA
- Medias Ajustadas
- Comparacion de medias Ajustadas.

Procedimiento en R para el analisis de covariancia.

Para el ejercicio, considere el archivo “crecimiento pijuayo”. Los tratamientos son dos localidades, la variable respuesta es altura y como variable externa el porcentaje de limo.

```
> datos <- read.table("crecimiento pijuayo.txt",header=T)
> attach(datos)
> modelo<-lm(altura~trat+limo,datos)
> anova(modelo)

Analysis of Variance Table

Response: altura
          Df Sum Sq Mean Sq F value   Pr(>F)
trat       1 16.576  16.576 12.8049 0.001335 **
limo       1  5.465   5.465  4.2216 0.049708 *
Residuals 27 34.952   1.295
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
> modelo$"coefficients"
(Intercept)      tratL2        limo
  9.0556366   1.3935427   0.1074507
> beta<-as.numeric(modelo$"coefficients"[3])
```

```
> beta
[1] 0.1074507
> cm <- cm.error(modelo)
```

El modelo encontrado por R, corresponde a un modelo lineal; esto significa que la estimacion del error es insesgado, puede ser utilizado para cualquier operacion de comparacion que se requiera. Sin embargo, la suma de cuadrados de tratamientos no esta ajustado por la variable X. Si se quiere comparar los promedios de tratamientos ajustado, necesita que se realice las operaciones apropiadas como se indica en el calculo de las medias ajustadas.

Homogeneidad de coeficientes de regresión.

Cuando el Diseño es completamente al Azar, se puede hallar la regresion de X en Y en cada tratamiento, y mediante una prueba de F, se puede probar si los coeficientes son homogeneos.

El procedimiento de ajuste es similar que en el proceso del ANCOVA.

1. Se calcula SC(x), SC(y) y SP(xy) para cada tratamiento.
2. Se halla la SC(regresion) y SC(residual) para cada tratamiento.
   SC(regresion)=[SP(xy)]² / SC(x) ; SC(residual) = SC(y) – SC(regresion)
3. Construir una tabla de estos resultados y halla ciertos totales y el ajuste.

Trat.	Gl.	SC(x)	SP(xy)	SC(y)	SC(residual)	Gl aj.
1	r_1-1					r_1-2
2						
...						
Totales	r. – t	Σ SC(x)	Σ SP(xy)	Σ SC(y)	Σ SC(residual)	r. – 2t

4. Se obtiene el; total de regresiones simple y la diferencia para la homogeneidad de regresiones:

   Total Residual: A = Σ SC(residual)
   Total de regresiones: B = Σ SC(y) - [Σ SP(xy)]² / Σ SC(x)
   La diferencia de homogenidad: B - A

5. Finalmente se realiza la prueba de F.

$$Fc = \frac{(B-A)/(t-1)}{A/(r.-2t)}$$

El valor de la tabla F 0.05 (t-1, r.-2t).

Se concluye la existencia de homogenidad cuando el valor Fc es menor al valor tabular.

Programa R para el analisis de homogeneidad de regresiones.

En el paquete “agricolae”, se tiene la funcion:

“reg.homog”

Con parametros:

Trat : grupos o tratamientos
Y : respuesta, variable dependiente
X : variable externa observada, independiente.

Para la prueba de esta funcion, utilice los siguiente datsos de ejemplo:

Caso que existe homogenidad:

Archivo: agr_homo.txt (datos de 4 tecnologias y el rendimiento en frijol)

En R, leer los datos:
> agricultor <- read.table("agr_homo.txt",header=TRUE)
> resultados<- reg.homog(agricultor$tecnologia, agricultor$produccion, agricultor$indice)

```
Prueba de Homogeneidad de regresiones

Total de regresiones simples:  4
Total de residuales         :  1490150
Diferencia para homogeneidad:  98847

Gl. homogeneidad regresiones:  3
Gl. Residuales              :  76
F.calculado                 :  1.680451
Pr > F                      :  0.1782828

Conclusión:  Existe homogeneidad de regresiones
```

Ejercicio: Realizar el analisis de homogenidad de regresion con los datos del archivo:

"regre_homo.txt"

el resultado debe ser:

```
F.calculado                 :  26.67623
Pr > F                      :  0.001033089

Conclusión:  no existe homogeneidad de regresiones
```

9. Experimentos con factoriales Cruzados y anidados (Jerárquicos).

Los factoriales son combinaciones de factores (nitrógeno, fósforo, variedades, sustancias, niveles de concentrado, etc.) para formar tratamientos, los cuales se aplican en los diseños experimentales (DCA, DBCA, DCL). La información obtenida de estos experimentos es amplia, ya que permiten comparar los niveles de cada factor entre sí y evaluar las interacciones que resulten como combinaciones de los factores, así como la comparación de niveles de un factor bajo un nivel de otro factor.

En un experimento con factoriales, si todos los niveles de un factor se combinan con todos los niveles de otro factor, entonces los factores son cruzados, si los niveles de un factor se combinan con ciertos niveles de otro factor se dice que estos factores están anidados.

Para tres factores se tiene los siguientes modelos:

Caso 1

a1				a2			
b1		b2		b1		b2	
c1	c2	c1	c2	c1	c2	C1	c2

Caso 2

a1				a2			
b1		b2		b3		b4	
c1	c2	c3	c4	c5	c6	c7	c8

Caso 3

a1				a2			
b1		b2		b3		b4	
c1	c2	c1	c2	c1	c2	c1	c2

Caso 4

a1				a2			
b1		b2		b1		b2	
c1	c2	c3	c4	c5	c6	c7	c8

Caso 1. Todos están cruzados
Caso 2. A cruzado, B y C anidado
Caso 3. A y C cruzados y B anidado
Caso 4. A y B Cruzado y C anidado.

<u>Tipos de factores</u>

Factores Cuantitativos.- Si sus niveles son cantidades cuantificables. Ejemplo. Niveles de Fósforo a 0.5%, 1% y 1.5%

Factores Cualitativos.- Si sus niveles no tienen orden natural y corresponden a clases o categorías. Ejemplo. Variedades de fríjol.

Ejemplo. Se tiene un factor conformado por 3 sustancias de crecimiento a 4 niveles de concentración aplicados para evaluar la propagación vegetativa de un cultivo sobre medios artificiales. La formación de callos se medirá a la cuarta semana.

El factor (A) sustancia de crecimiento con niveles:

a_1 : Ácido Indolacético (AIA)
a_2 : Cinetina (C)
a_3 : Ácido Naftaleno acético (ANA)

El factor (B) concentración con niveles:

b_1 : 0.0
b_2 : 0.1 μM
b_3 : 1.0 μM
b_4 :10.0 μM

Al combinar ambos factores A y B se tienen 3x4 = 12 tratamientos para ser evaluados.

Los factores se identifican con letras mayúsculas y los niveles con letras minúsculas, por ejemplo:

Factor sustancia = A con niveles a_1, a_2, a_3

Factor concentración = B con niveles b_1, b_2, b_3, b_4

La combinación resultante: a_1b_1, a_1b_2, a_1b_3,..., a_3b_4

Estos tratamientos son:

a_1b_1 = 0.0 concentración de AIA
a_1b_2 = 0.1 μM concentración de AIA
....
....
a_3b_4 = 10 μM de concentración de ANA

Si cada tratamiento se aplica a 4 unidades experimentales, se requiere 48 u.e. para realizar el experimento.

Los factoriales son expresados mediante la siguiente notación:

2A2B = 2x2 = 2^2: 2 niveles de A por 2 niveles de B.
2A3B = 2x3: 2 niveles de A por 3 niveles de B.
2A2B2C = 2x2x2 = 2^3: 3 factores a 2 niveles cada uno.
2A3B3C = $2x3^2$: 2 niveles de A por 3 niveles de B y 3 niveles de C.

Formación de factoriales

En la formación de factoriales, se debe tener presente lo siguiente:

1. - Que factores deben incluirse.
2. - Que factores son fijos (modelo I) y que factores son al Azar (modelo II).
3.- Cuantos niveles por factor
4. - Si son factores cuantitativos, cual debe ser el espaciamiento entre los niveles del factor. Por ejemplo: 0%, 5% y 10% de nitrógeno, significa igual espaciamiento.

Ventajas y desventajas en experimentos con factoriales

Los experimentos con factoriales tienen las siguientes ventajas:

1. - Permiten el estudio de los niveles de cada factor y las interacciones entre ellos.

2. - Permiten el estudio de los niveles de un factor en la combinación de un sólo nivel de otro factor (estudio de efectos simples).

3.- Todas las unidades experimentales intervienen en el estudio de todos los efectos del factor (principales e interacción).

Desventajas:

1.- El número de unidades experimentales utilizadas es mayor que en experimentos simples y es más difícil contar con un número suficiente de unidades que requiere el experimento.

2.- El análisis se complica, a medida que el número de factores y niveles aumenta.

3.- Algunas combinaciones pueda que no sean de importancia, pero deben incluirse para completar el factorial, esto obliga a usar más unidades experimentales.

Interacción de factores.

La interacción es una fuente muy importante en el estudio de los experimentos con factoriales, permite evaluar las diferencias de niveles de un factor en la combinación de un tercero.

Las interacciones se presentan solo en los casos que ambos son factores cruzado o cuando un factor Anidado antecede a un cruzado.

Así, para los casos presentados, las interacciones que pueden estudiarse son:

Caso 1: AB, AC, BC y ABC
Caso 2: Ninguna interacción.
Caso 3: AC y B(A) con C
Caso 4: AB

Fuentes de variación para los casos presentados:

Caso 1.

Fuentes	Gl
A	a-1
B	b-1
C	c-1
AB	(a-1)(b-1)
AC	(a-1)(c-1)
BC	(b-1)(c-1)
ABC	(a-1)(b-1)(c-1)
Error	(r-1)abc
Total	rabc -1

Caso2

Fuentes	Gl
A	a-1
B(A)	(b-1)a
C(AB)	(c-1)ab
Error	(r-1)abc
Total	rabc – 1

Caso 3

Fuentes	Gl
A	a-1
B(A)	(b-1)a
C	c-1
AC	(a-1)(c-1)
B(A)C	(b-1) a (c-1)
Error	(r-1)abc
Total	rabc-1

Caso4

Fuentes	Gl
A	a-1
B	b-1
C(AB)	(c-1)ab
AB	(a-1)(b-1)
Error	(r-1)abc
Total	rabc-1

Para un correcto análisis de variancia, es necesario determinar los componentes de variancia de cada fuente de variación para conocer cual es el denominador apropiado para una prueba de F.

Componentes de los esperados cuadrados medios

Para hallar estas componentes, es necesario precisar que factor es fijo y cual al azar.

Factor fijo.- Cuando todos los niveles del factor están en el experimento.

Factor al azar.- Cuando un numero pequeño de niveles del factor están en el experimento.

A estas dos respuestas, se debe agregar si el factor es cruzado o anidado.

Regla para hallar las componentes de variancia

1. Halle todas las fuentes de variación en función del anidamiento y el cruce de factores
2. Determine los factores fijo y al azar.
3. Construya una matriz, en la primera columna las fuentes de variación, y en la primera fila los factores, se incluye las repeticiones como un factor al azar.
4. Se llena las celdas con el siguiente criterio:
    - Primero el concepto fijo (valor 0) y al azar (valor 1) estos valores se llenan en cada columna dependiendo si la fuente contiene el factor.

- Segundo el concepto de anidamiento (valor 1). Este valor se coloca en la celda si el factor de la columna aparece como anidador en la fuente.
- Completar las celdas de cada columna con el respectivo valor de niveles que tiene el factor.
- Todas las celdas de repeticiones con el numero de repeticiones.
- Todas las celdas de la fila del error con el valor 1.
- En la columna de variancia coloque el símbolo de la variancia de la fuente tratada.

5. Una vez completado el cuadro, se procede a hallar las componentes de variancia uno por uno, siguiendo el criterio:
   - Cada fuente tiene una letra o más, esta identificación sirve para ocultar las columnas de las letras que se indica en la fuente .
   - Se multiplica los otros coeficientes solo las que corresponda, empezando de abajo hacia arriba.

Para ilustración de estas reglas, se considera el caso 3 presentado, con factores A al azar con 2 niveles, B fijo con 3 niveles y C fijo con 2 niveles con 4 repeticiones.

Caso 3													
	a1						A2						
	b1		b2		b3		B4		b5		b6		
	c1	c2	c1	c2	c1	c2	C1	c2	c1	c2	C1	c2	

Fuentes	A	B	C	R	Variancia	E[CM]
A	1	3	2	4	Va	Ve + 24 Va
B(A)	1	0	2	4	Vab	Ve + 8 Vab
C	2	3	0	4	Vc	Ve + 12 Vac + 24 Vc
AC	1	3	0	4	Vac	Ve + 12 Vac
B(A)C	1	0	0	4	Vabc	Ve + 4 Vabc
Error	1	1	1	1	Ve	Ve

En la columna A:

- Coloque 1 en el intercepto con A y AC. (A es un factor al azar)
- Coloque 1 en el Intercepto B(A) y B(A)C. (A es el anidador)
- Coloque 2 en la celda restante.

En la columna B:

- Coloque 0 en el intercepto con B(A) y B(A)C. (B es fijo)
- Coloque 3 en la celdas restantes menos el error.

En la columna C:

- Coloque 0 en el intercepto con C, AC y B(A)C. (C es fijo)
- Coloque 2 en las celdas restantes menos el error.

En la columna R todas las celdas son 4 menos en el error
En la fila del error todas las celdas son 1
En la columna de variancia se coloca el símbolo de la variancia de la fuente.

Hallando las componentes de variancia.

Para la fuente A:

- Se oculta la columna A.
- Se multiplica las celdas solo donde se tiene A en la fuente, así:
  1.1.1.Ve = Ve
  0.0.4.Vabc = 0
  3.0.4.Vac = 0
  0.2.4.Vab = 0
  3.2.4.Va = 24 Va
- Resulta Ve + 24 Va

Para la fuente B(A)

- Se oculta las columnas A y B.
- Se multiplica las celdas solo donde se tiene B(A) en la fuente, así:
  1.1.Ve = Ve
  0.4.Vabc = 0
  2.4.Vab = 8 Vab
- Resulta Ve + 8 Vab

Para la fuente C:

- Se oculta la columna C.
- Se multiplica las celdas solo donde se tiene C en la fuente, así:
  1.1.1.Ve = Ve
  1.0.4.Vabc = 0
  1.3.4.Vac = 12 Vac
  2.3.4.Va = 24 Va
- Resulta Ve + 12 Vac + 24 Vc

Fuente AC:

- Se oculta las columnas A y C.
- Se multiplica las celdas solo donde se tiene A y C juntos, así:
  1.1.Ve = Ve
  0.4.Vabc = 0
  3.4.Vac = 12 Vac
- Resulta Ve + 12 Vac

Fuente B(A)C:

- Se oculta las columnas A, B y C.
- Se multiplica las celdas solo donde se tiene B(A) y C juntos, así:
  1.Ve = Ve
  4.Vabc =4 Vabc
- Resulta Ve + 4 Vabc

Para el error, el estimado de la variancia es Ve.

Según este análisis, las relaciones para el F calculado son las siguientes:

Fuentes	Gl	F- calculado
A	1	CM(A)/CM(error)
B(A)	4	CM(B(A))/CM(error)
C	1	CM(C)/CM(AC)
AC	1	CM(AC)/CM(error)
B(A)C	4	CM(B(A)C)/CM(error)
Error	36	
Total	47	

10. <u>Diseño en Parcela Dividida</u> (Split plot)

El diseño consiste en dividir las parcelas en subparcelas y aplicar un grupo de tratamientos en las parcelas y otro grupo de tratamientos en las subparcelas. La asignación de los tratamientos en las parcelas (unidades experimentales) se realiza de acuerdo al criterio del diseño aplicado (DCA, DBCA, DCL) y en las subparcelas los tratamientos se aplican al azar completamente.

Por ejemplo: Un diseño completamente al azar en parcelas divididas con un factor A de 3 variedades de un cultivo en parcelas y un factor B de 2 niveles de fertilización en subparcelas, con 4 repeticiones:

Para este ensayo se requiere 12 parcelas grandes y dividirla cada una en dos subparcelas.

El croquis en campo podría ser:

<table>
<tr><td colspan="2">a1</td><td colspan="2">a2</td><td colspan="2">a1</td></tr>
<tr><td>b1</td><td>b2</td><td>b2</td><td>b1</td><td>b2</td><td>b1</td></tr>
</table>

<table>
<tr><td colspan="2">a3</td><td colspan="2">a1</td><td colspan="2">a2</td></tr>
<tr><td>b2</td><td>b1</td><td>b2</td><td>b1</td><td>b1</td><td>b2</td></tr>
</table>

<table>
<tr><td colspan="2">a2</td><td colspan="2">a3</td><td colspan="2">a2</td></tr>
<tr><td>b1</td><td>b2</td><td>b1</td><td>b2</td><td>b2</td><td>b1</td></tr>
</table>

<table>
<tr><td colspan="2">a3</td><td colspan="2">a3</td><td colspan="2">a1</td></tr>
<tr><td>b2</td><td>b1</td><td>b2</td><td>b1</td><td>b1</td><td>b2</td></tr>
</table>

Observe que los tratamientos a_1, a_2 y a_3 están distribuidos en las parcelas (unidades experimentales) según el diseño, en este caso aleatoriamente, cada tratamiento A_i repetidas cuatro veces. El factor B con los niveles b_1 y b_2 son aplicados aleatoriamente en las subparcelas de cada parcela.

Este diseño tiene dos tipos de errores no controlables, un error generado por las diferencias entre las parcelas en el cual se aplicó el mismo tratamiento A_i, y un error de las diferencias de las subparcelas en el cuál figura una combinación de A y de B, por ejemplo las parcelas en las que se aplicaron el nivel A_2 y de estas, las subparcelas que recibieron el nivel b_3.

También se observa que las parcelas son como pequeños bloques, porque en cada subparcela se aplica una combinación de A y B que son los tratamientos de la combinación A y B, estos bloques no son completos, porque no están todos los tratamientos, sin embargo se afirma que hay un efecto de estos bloques incompletos que está mezclado con el efecto de los tratamientos de A, esto significa que los efectos principales de A se han confundido con el efecto de los bloques incompletos.

Error(a) y Error(b)

El error en parcelas es el Error(a) y el error en subparcelas es el Error(b). El CM(error(a)) se representa por E_a y el CM(error(b)) por E_b.

Por lo general E_a es superior a E_b, esto se debe a que las observaciones en las subparcelas de la misma parcela tienden a correlacionarse positivamente. E_a no puede ser menor que E_b, excepto por el azar y si esto sucede, se puede consideran como estimadores de σ^2 una combinación de los dos errores, así:

$$\hat{\sigma}^2 = \frac{SC(error(a)) + SC(error(b))}{Gl(error(a)) + Gl(error(b))}$$

El CM(error(a)) representado por E_a, se utiliza para relacionar los Cuadrados Medios de las fuentes que están a nivel de parcelas y el CM(error(b)) representado por E_b, para relacionar los cuadrados medios de las fuentes que están a nivel de subparcelas, sólo para el caso de factores fijo. Si los factores son fijos y al azar, mezcla de ambos, la relación de los cuadrados medios estarán sujeta a las reglas de los factoriales cuando los factores son fijos y al azar.

Como se tiene 2 errores, también se tiene dos coeficientes de variación, dado por:

$$CV(a) = \frac{\sqrt{Ea/b}}{\bar{Y}} x100\%, \quad CV(b) = \frac{\sqrt{Eb}}{\bar{Y}} x100\%$$

Si el error(b) es más grande que el error(a), el coeficiente de variación debe ser expresado por:

$$CV = \frac{\sqrt{\hat{\sigma}^2}}{\bar{Y}} x100\%$$

Algunos criterios para asignar los factores A y B.

El factor A siempre se aplica a parcelas grandes o en gran escala como en ensayos industriales y de laboratorio, y el factor B el que se suministra a pequeña escala.

En el caso de poder elegir, el factor más importante debe ser el que se aplique en subparcelas, es decir factor B.

Debe tener en cuenta que el efecto del factor A esta confundido con las diferencias de los bloques incompletos formado por el factor B.

Características:

El control local se realiza en las parcelas, estas deben ser tratadas de acuerdo al tipo de diseño (Completo al azar, Bloques, Latino, etc.), Estas deben ser divididas en subparcelas, un numero igual a los niveles de B.

La aleatorización debe realizarse en dos fases, en las parcelas grandes con el factor A de acuerdo al diseño utilizado, y en las subparcelas el factor B se aplica completamente al azar.

Respecto a las repeticiones, las parcelas grandes constituyen las repeticiones del experimento, y deben ser un numero tal que los grados de libertad del error(a) tengan un valor considerable, por ejemplo no menor de 10.

Las fuentes de variación generadas dependerán del diseño aplicado (DCA, Bloque, DCL u otro)

Así por ejemplo DCA.

Fuentes	G.L.
A	a-1
Error(a)	a(r-1)
B	b-1
AB	(a-1)(b-1)
Error(b)	a(b-1)(r-1)
Total	abr-1

En el caso de Bloques

Fuentes	G.L.
Bloques	r-1
A	a-1
Error(a)	(a-1)(r-1)
B	b-1
AB	(a-1)(b-1)
Error(b)	a(b-1)(r-1)
Total	abr-1

En el programa R, los modelos podrian ser escritos como:

```
> Modelo.DCA <- aov( y ~ A + Error(rep/A) + B + A:B )
> Modelo.DBCA <- aov( y ~ bloques + A + Error(bloques:A) + B + A:B )
```

Como se puede observar, los cambios en las fuentes de variación solo se da a nivel de parcelas, no a nivel de subparcela.

11. Diseño con Bloques Divididos.

En experimento con dos factores, en muchos casos no es posible combinar para formar tratamientos y aplicar en la unidad experimental, una posibilidad es aplicar los niveles de los factores en franjas. Este experimento es apropiado en áreas grandes.

La precisión es mayor en la medición de la interacción que en los factores de estudio.

Para comprender mejor el diseño, considere dos factores A de 4 niveles y B de tres niveles aplicados en 3 repeticiones.

En cada bloque, sé aleatoriza por columna los niveles de A y por fila los niveles de B, en cada parcela se dispondrán el tratamiento que combina la fila con la columna.

Bloque I				
	a4	a2	a3	a1
b2				
b3				
b1				

Bloque II				
	a2	a1	a3	a4
b2				
b1				
b3				

Bloque III				
	a4	a2	a4	a3
b3				
b2				
b1				

El análisis y el diseño se ajusta a la siguiente distribución de los grados de libertad.

Fuentes	Gl	Ejemplo
Bloques	r-1	2
A	a-1	3
Error (a)	(r-1)(b-1)	6
Total de parcelas en A	ra – 1	11
B	b-1	2
Error (b)	(r-1)(b-1)	4
Total de parcelas en B	b(r-1)	6
AB	(a-1)(b-1)	6
Error (c)	(r-1)(a-1)(b-1)	12
Total de subparcelas	abr – 1	35

Para ganar precisión en el factor B, los niveles de este factor pueden ser aleatorizado formando un cuadrado latino a través de las repeticiones, implica que el numero de niveles de B debe ser igual al número de repeticiones.

Así por ejemplo:

Bloque I				
	A4	a2	a3	a1
b2				
b3				
b1				

Bloque II				
	a2	a1	a3	A4
b3				
b1				
b2				

Bloque III				
	A4	a2	a4	a3
b1				
b2				
b3				

El análisis y el diseño se ajustan a la siguiente distribución de los grados de libertad.

Fuentes	Gl	Ejemplo
Bloques	r-1	2
A	a-1	3
Error (a)	(r-1)(b-1)	6
Total de parcelas en A	ra – 1	11
Hileras	b-1	2
B	b-1	2
Error (b)	(b-1)(b-2)	2
AB	(a-1)(b-1)	6
Error (c)	(r-1)(a-1)(b-1)	12
Total de subparcelas	abr – 1	35

Este diseño respecto del anterior, solo se tiene una ventaja que corresponde al factor B que es medido con mas eficiencia, el error (b) disminuye en grados de libertad y en suma de cuadrados, que favorece al factor B.

En ambos diseños, las pruebas se realizan usando como denominador a su respectivo error, así para el factor A es el error (a), para el Factor B el error (b) y para la interacción el error (c).

El analisis de este diseño se muestra en las aplicaciones, con su programa en R.

12 . Experimentos factorials sin repetición.

En la mayoría de los investigadores estan interesados en investigar los efectos de una variable de las variaciones conjuntas de muchos factores, sin embargo a medida que se adiciona mas factores, el experimento se hace muy grande imposible de conducirlo, en estos casos, lo recomendable es confundir efectos de interacción con el efecto de los bloques, este procedimiento reduce el numero de unidades experimentales.

Principio de la Confusión.

Suponga que el factorial 2x2x2 se planea en 4 bloques, esto obligaría disponer de 32 unidades experimentales.

Si se examina los efectos del factorial 2x2x2 para ser evaluado por contrastes, se obtiene:

Fuente	(1)	c	b	cb	a	ac	ab	Abc
A	-1	-1	-1	-1	1	1	1	1
B	-1	-1	1	1	-1	-1	1	1
C	-1	1	-1	1	-1	1	-1	1
AB	1	1	-1	-1	-1	-1	1	1
AC	1	-1	1	-1	-1	1	-1	1
BC	1	-1	-1	1	1	-1	-1	1
ABC	-1	1	1	-1	1	-1	-1	1

Estos significa, para confundir el efecto de la interacción ABC, se puede formar semibloques, de tal forma que se contraste tal efecto; asi:

I	abc	a	b	c
II	(1)	ac	ab	bc

En cada bloque se aleatoriza.

El investigado puede confundir las interacciones de menos interes. Si en un experimento se confunde un solo tipo de efecto se dice que es una confusión completa; por el contrario si se confunde mas de una interacción se indica que la confusión es parcial.

No existe restricción en esgoger las interacciones que se desea confundir.

Procedimiento R para realizar diseños confundidos.

Requiere la biblioteca: **conf.design**

Valido solo para factorial $\mathbf{2^n}$, $\mathbf{3^n}$, $\mathbf{4^n}$, etc

Confundir la interaccion AB en el factorial 2^2.

D22 <- conf.design(c(1,1), p=2, treatment.names=LETTERS[1:2])

```
> D22
  Blocks A B
1      0 0 0
2      0 1 1
3      1 1 0
4      1 0 1
```

Confundir la interaccion ABC en el factorial 2^3.

D222 <- conf.design(c(1,1,1), p=2, treatment.names=LETTERS[1:3])

```
> D222
  Blocks A B C
1      0 0 0 0
2      0 1 1 0
3      0 1 0 1
4      0 0 1 1
5      1 1 0 0
6      1 0 1 0
7      1 0 0 1
8      1 1 1 1
```

Confundir la interaccion AB en el factorial 2^3.

Dab <- conf.design(c(1,1,0), p=2, treatment.names=LETTERS[1:3])

```
> Dab
  Blocks A B C
1      0 0 0 0
2      0 1 1 0
3      0 0 0 1
4      0 1 1 1
5      1 1 0 0
6      1 0 1 0
7      1 1 0 1
8      1 0 1 1
```

Análisis Estadístico de experimentos confundidos.

Las fuentes de variación se ven restringidas por las interacciones confundidas. Por ejemplo si se realiza una confusión completa para la interacción ABC en 6 semibloques, la distribución de las fuentes de variación y los grados de libertad son los siguientes:

Fuentes	Gl
Repetición	5
A	1
B	1
C	1
AB	1
AC	1
BC	1
Error	12

La determinación de las sumas de cuadrados son calculados en forma usual y las pruebas de F en el análisis de variancia es determinado usando el error como fuente del denominador.

Considere el siguiente ejercicio para comprender el analisis:

	abc	a	b	c
I	30	16	24	19
III	23	16	28	16
V	25	19	20	16

	ab	ac	bc	(1)
II	28	16	27	8
IV	18	25	16	10
VI	23	22	17	18

```
Response: respuesta
          Df  Sum Sq Mean Sq F value  Pr(>F)
bloque     5  53.000  10.600  0.4837 0.78216
a          1  73.500  73.500  3.3536 0.09198 .
b          1 253.500 253.500 11.5665 0.00526 **
c          1  24.000  24.000  1.0951 0.31597
a:b        1   6.000   6.000  0.2738 0.61034
a:c        1  13.500  13.500  0.6160 0.44777
b:c        1  37.500  37.500  1.7110 0.21535
Residuals 12 263.000  21.917
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

Promedio = 20
CV = 23.4 %

Programa en R. Analisis de experimento confundido **2^3 (A*B*C)**

```
# cargar las bibliotecas requeridas
library(combinat)
library(agricolae)
#
# Leer los datos del archivo factorial2x2x2.txt
# factores : bloque a b c
# variable : respuesta
# Diseño confundido 2x2x2, la interaccion A*B*C es confundida
```

```
# por la formacion de bloques incompletos
#
rm(list=ls())
datos <- read.table("factorial2x2x2.txt",header=TRUE)
#
attach(datos)
conjunto <-data.frame(bloque=factor(bloque),a=factor(a),b=factor(b),
c=factor(c),respuesta=respuesta)
modelo <-aov(respuesta ~ bloque + a*b*c -a:b:c,conjunto)
print(modelo)
print(anova(modelo))

# Utilizar el procedimiento agricolae

gl<-df.residual(modelo) # extrae los grados del libertad delr error
error<-deviance(modelo)/gl # extrae el CM(error)
# realiza las comparaciones multiples LSD.
compara.a<-LSD.test(respuesta,a,gl,error) # para A
compara.b<-LSD.test (respuesta,b,gl,error) # para B
compara.c<-LSD.test (respuesta,c,gl,error) # para C
promedio<-mean(respuesta)
cv<-sqrt(error)*100/promedio
```

13. Experimentos Repetidos en espacio y tiempo.

En muchas ocasiones un experimento es repetido en mas de una vez en el mismo lugar, en diferentes lugar y en otras épocas. Generalmente es una política de una institución sobre la investigación realizada, mas que el interés propio de un investigador. Estas prácticas involucran personal, tiempo y terrenos disponibles para ello, además del costo que demanda. Sus resultados son a mediano o largo plazo. Los experimentos repetidos son consistentes e involucra objetivos mayores propios de la Institución que realiza la investigación.

Cada experimento repetido, debe seguir su propia randomización, lo único que debe ser común, es el mismo diseño con un numero de tratamientos iguales y repeticiones.

En Cultivos anuales, cada periodo es una nueva evaluación y las plantas cambian; entonces, las fuentes de variación del analisis combinado es el siguiente:

Lugar 1.
6 bloques y 4 tratamientos

En un periodo:

Fuentes	Gl
Bloques	5
Tratamientos	3
Error	15
Total	23

El combinado de dos periodos en el lugar 1 será:

Fuentes	Gl
Periodo	1
Bloques(Periodo)	5 (constituye el error (a))
Tratamientos	3
Tratamiento*periodo	3
Error	15 (constituye el error (b))
Total	23

Si no existe bloques, esta fuente suma el error en ambos casos.

Si el experimento se realiza en otra localidad bajo las mismas condiciones, el combinado será:

Fuentes	Gl
Lugar	1
Bloque(lugar)	10 (constituye un error (a))
Periodo	1
Lugar*periodo	1
Periodo*bloque(lugar)	10 (constituye un error (b))
Tratamiento	3
Tratamiento *periodo	3
Tratamiento*lugar	3
Tratam.*periodo*lugar	3
Error	60 (constituye un error (c))
Total	95

Si no existen bloques, esto se suma el error y se tendría un solo error para todas las pruebas siempre y cuando los lugares y periodos se consideren fijos.

En experimentos con cultivos perennes durante varios años, en donde las plantas de cada parcela son las que se cosechan en años sucesivos y los tratamientos no cambian de posición a diferencia de cultivos anuales, la continuidad de los tratamientos en las mismas parcelas durante todos los años, permite considerar la cosecha de una parcela como una subparcela y el total de cosechas de todos los años como una parcela y así se analiza el experimento como un modelo de parcelas divididas. En este tipo de experimentos, no es necesario probar homogeneidad de variancia del error de todos los años.

Por ejemplo, se prueba en una localidad 4 tratamientos bajo el Diseño de Bloques (6 bloques) y en el segundo año se hace otra evaluación en el mismo sitio (en las mismas parcelas).

Localidad 1.

En un periodo:

Fuentes	Gl
Bloques	5
Tratamientos	3
Error	15
Total	23

El combinado con el segundo periodo resulta:

Fuentes	Gl
Tratamientos	3
Bloques	5
Tratamientos*bloque	15 (constituye un error (a))
Periodo	1
Bloque *periodo	5
Tratamiento*periodo	3
Error	15
Total	47

Si el mismo diseño se realizo en otro lugar en los mismos periodos, entonces el combinado con localidad seria:

Fuentes	Gl
Lugar	1
Bloque(lugar)	10 (constituye un error (a))
Tratamiento	3
Tratamiento*lugar	3
Tratamiento*bloque(lugar)	15 (constituye un error (b))
Periodo	1
Lugar *periodo	5
Periodo*bloque(lugar)	10
Lugar*tratamiento*periodo	3
Error	15 (constituye un error (c))
Total	95

14. Experimento en campo de agricultores.

La Investigación agrícola, se realiza por lo general en estaciones experimentales, porque se dispone de condiciones apropiadas para ello. Las tecnologías desarrolladas en las estaciones no siempre son utilizadas directamente por los agricultores, las cuales se deben probar en campo de agricultores. La investigación en campos de agricultores son de dos tipos: La investigación para generar nuevas tecnologías y la investigación para verificación de tecnologías. En ambos experimentos, los ensayos deben ser conducidos en mas de dos agricultores simultáneamente, considerando una fuente de variación el efecto del campo del agricultor.

Uno de los problemas en estos experimentos, es la selección del sitio de prueba, que debe ser lo mejor en términos de homogeneidad (pendiente, suelo, agua, clima, etc.).

Si se va ha realizar experimentos para generar tecnología, la selección del sitio debe ser primordial, y se sugiere lo siguiente:

Paso 1. Especifique la prueba deseada para tratar el medio ambiente en términos de las características físicas y biológicas de clima, topografía, suelo, régimen de agua. y así sucesivamente

Paso 2. Clasifique cada uno de las características del medio ambiente, especificadas según:

El tamaño relativo de área homogénea de una característica dada. Por ejemplo áreas con el mismo clima serán más grande que aquéllas que tienen el mismo paisaje o régimen de agua.

La disponibilidad de información existente o la obtención de la información de las características deseadas. Por ejemplo, los datos de clima está normalmente disponible que la información sobre paisaje, régimen de agua. El último normalmente se obtiene a través de las visitas del campo por el investigador.

Paso 3. Seleccione una área grande que satisfaga las características ambientales con rasgos homogéneos, por ejemplo clima y suelo, dentro de esta área grande seleccione áreas mas pequeñas, utilice un plano para describir la topografía, régimen de agua, etc.

Paso 4. Finalmente, de cada agricultor, debe disponer de áreas lo mas grande posible para que se adecuen al experimento y con mínima heterogeneidad del suelo. Si no es posible, en un agricultor, seleccione un grupo pequeño de áreas del agricultor.

EL Diseño y análisis.

El tamaño del experimento debe ser pequeño, esto referido al numero de tratamientos y repeticiones, para estos casos se podría utilizar experimentos fraccionados, si hay pocas unidades experimentales. El mismo ensayo que se aplique en un agricultor, se debe aplicar a los otros agricultores, para hacer un análisis conjunto.

Fuentes.	Grados de libertad
Agricultores	a-1
Repetición (Agricultores)	(r-1)a
Tratamientos	t-1
Tratamientos x Agricultor	(t-1)(a-1)
Error Experimental	(r-1)(t-1)a
Total	rta – 1

Cuando se utiliza tratamientos testigos, se puede utilizar contrastes ortogonales.

Por ejemplo:

4 agricultores, 4 tratamientos mas un tratamiento testigo (t5)

Fuentes.	Grados de libertad
Agricultores	3
Repetición (Agricultores)	3(r-1)
Tratamientos	4
Tratamientos x Agricultor	(12)
(t5 vs otros) x agricultor	3
(t1, t2 vs t3, t4) x agricultor	3
(t1 vs t2) x agricultor	3
(t3 vs t4) x agricultor	3
Error Experimental	16(r-1)
Total	20r – 1

En experimentos para verificación de tecnologías, se requiere mas agricultores que para prueba de generar tecnologías, las pruebas están basadas por comparaciones con la practica del agricultor. En experimentos para generar tecnologías, se utiliza tratamientos testigos con nivel cero, en este caso, el control es la practica del agricultor, el cual realiza algunas actividades como por ejemplo aplicar fertilizantes y realizar control de insectos que se esta recomendando.

En experimento para generar tecnología, todas las pruebas, excepto las practicas, son realizas al nivel optimo, en cambio en las pruebas de verificación, son mantenidas a nivel del agricultor.

Diseño y Análisis.

En general se debe aplicar un factorial 2^k , son k factores o componentes, cada factor con dos niveles (Nueva tecnología, tecnología del agricultor). Si son 4 componentes la nueva tecnología, entonces el factorial es 2^4 = 16 tratamientos. En estos casos, se puede fraccionar el factorial a la mitad para tener solo 8 tratamientos, sacrificando las interacciones de mayor orden.

En todos estos experimentos, se requiere repeticiones en cada campo del agricultor. En otras situaciones, donde no es posible tener repeticiones, se puede utilizar algunas técnicas estadísticas, utilizando índices ambientales, permita de alguna forma buscar el mejor paquete tecnológico que se pueda utilizar el agricultor. Para este caso, se plantea la siguiente metodología:

Utilizar pocos tratamientos y muchos lugares o campos de agricultores, que represente a la diversidad de agricultores en el estudio.

Definir el tamaño de la parcela del agricultor, que debe ser igual al tamaño de todos los agricultores.

En cada sitio (agricultor), aplicar todos los tratamientos que se quiere probar. Como si cada agricultor fuese un bloque y el experimento un diseño en bloques.

Terminado el experimento, recolectar la información en una matriz de información clasificado por agricultor y tratamiento.

Para el análisis de los datos, estime una ecuación de regresión lineal simple por cada tratamiento.

$Y_i = a + b X_i$

Y_i = Respuesta de la parcela.
X_i = Indice ambiental, para este caso, considere el promedio de respuesta de cada agricultor, que representa el índice del lugar.

En base a estas respuestas, realice su primera selección, en base a los tratamientos mas estables.

Tratamientos en las cuales se tienen promedios altos, coeficiente b = 1 y menor cuadrado medio de la regresion, es decir estadísticamente no debe ser distinto de 1. Esto significa que los cambios en una unidad producida en promedio de los agricultores respecto de otro, debe ser la misma, es decir tratamientos estables, sin cambios entre agricultores.

Ho b = 1
H1 b $\neq$ 1

$$Tc = \frac{b-1}{\sqrt{\frac{CMerror\,Regresion}{\sum x_i^2}}}$$

Mediante pruebas de t se decide la aceptación o rechazo de las hipótesis planteadas..

b: Coeficiente de regresión estimado
ΣX^2: Suma de cuadrados del indice ambiental corregido

El valor critico se localiza en la tabla de t-student con dos colas y grados de libertad (Nro de ambientes – 2)

Para seleccionar el tratamiento, se utilizar los ambientes o agricultores en donde se ha logrado respuestas superiores al promedio general.

Con este nuevo conjunto de datos, puede proceder a hacer una selección mas especifica o comparar los tratamientos, hallando los intervalos de confianza para cada tratamientos a diferente nivel de probabilidad.

$$I.C. = \overline{X} \pm t\alpha \frac{S}{\sqrt{n}}$$

I.C.: Intervalo de confianza.
S: desviación estándar(solo de los ambientes buenos).
n: número de ambientes buenos.

Mediante una gráfica de los intervalos de confianza, se puede decidir si un tratamiento de los seleccionados es diferente o no, observado la superposición de areas de los intervalos de confianza.

Para claridad de la metodología, se acompaña un ejemplo al final del documento.

15. Aplicaciones

Analisis del Diseño de bloques con unidad Faltante.

Sustitución de sombras tradicionales con maderables en cacaotales establecidos. http://computo.catie.ac.cr/~gtz/cviejo.html
Objetivo

Evaluar el comportamiento de tres especies maderables y una leguminosa como sombra para un cacaotal establecido, así como su efecto sobre la producción de cacao. Además evaluar la factibilidad técnica y económica de un cambio de sombra en cacaotales establecidos.

- Fecha de inicio: 1989.
- Número de ensayos: 1.
- Ubicación: Guabito, Nuevo Paraíso y Finca 51, Changuinola, Panamá.
- Diseño experimental: Bloques completos al azar con cuatro tratamientos y seis repeticiones.
- Tratamientos: Uno por especie evaluada, Laurel (Cordia alliodora), Framiré (Terminalia ivorensis), Roble de sabana (Tabebuia rosea) y guaba (Inga edulis).

Variables: Diámetro a la altura de pecho (DAP) y altura total para árboles maderables.
DAP promedio (cm) por especie y sitio experimental (edad=4.4 años).

SITIO	LAUREL	GUABA	ROBLE	TERMINALIA
1	18.4	14.3	21	20.9
2	19.1	12.9	19.7	18.2
3	.	15	13.2	19.2
4	14.4	14.6	13.2	21.7
5	12.9	14.6	16.8	15.7
6	14.4	12.4	14	18.6

Indices de rentabilidad por especie de sombra maderable.

	Planicie		Colinas	
Especies	Razón B/C	US$ por jornal	Razón B/C	US$ por jornal
Laurel	5.3	41	5.1	40
Terminalia	4.4	34	5.1	41
Roble de sabana	4.4	33	4.6	36

Comentario

La guaba fue la peor especie para sustituir sombra en cacaotales establecidos. Su crecimiento y sobrevivencia fueron pobres en cacaotales densos. Desde el punto de vista de sombra para el cacao, el roble, el laurel y la terminalia presentan limitaciones dado que pierden la totalidad del follaje durante la época seca, periodo en que el cacao requiere más sombra. Los costos de renovación de sombra con especies maderables o leguminosas alcanzan entre 750 y 950 US$ en los primeros cuatro años. Este cambio de sombra puede generar en el año 15 beneficios por venta de madera entre 4500 y 5500 US$/ha. El cambio de sombra en cacaotales maduros con maderables efectuado bajo las condiciones del ensayo es una inversión rentable.

Estimación de la unidad Faltante.

Por Yates:

(6*47.4+4*79.2-375.2)/[(6-1)(4-1)] = 15.0667

R-

Fuentes	Gl	SC	CM	Fc
Sitios	5	43.9511	8.79022	1.71
Especies	3	79.6844	26.5615	5.16
Error	14	72.0306	5.14504	
Total	22	195.666		

Programa R para el analisis de las especies maderables.

```
# cargar las bibliotecas requeridas
library(combinat)
library(agricolae)
#
# Leer los datos del archivo maderable.txt
# factores : sitio especie
# variable : diametro
# Diseño bloques completos al azar con un dato faltante

datos <- read.table("maderable.txt",header=TRUE)
#
attach(datos)
conjunto <-data.frame(sitio=factor(sitio),especie=factor(especie),diametro=diametro)
modelo <-aov(diametro ~ sitio + especie,conjunto)
print(modelo)
print(anova(modelo))

# Utilizar el procedimiento taller

gl<-df.residual(modelo) # extrae los grados del libertad delr error
error<-deviance(modelo)/gl # extrae el CM(error)
compara<-LSD.test(diametro,especie,gl,error) # realiza las comparaciones multiples LSD.
```

Análisis del Diseño cuadro latino.

"Evaluación del sistema de riego por exudación utilizando cuatro variedades de melón, bajo modalidad de siembra, SIMPLE HILERA.". Se desea probar el comportamiento de tres variedades híbridas de melón y uno estándar. (Tesis).- autor Alberto Ángeles L.

Variedades: V1 : Híbrido Mission
V2 : Híbrido Mark.
V3 : Híbrido Topfligth.
V4 : Híbrido Hales Best Jumbo.

Datos: Rendimiento en Kgs. por parcela.

	C1	C2	C3	C4
F1	45	50	43	35
F2	29	53	41	63
F3	37	41	41	63
F4	38	40	35	41

	C1	C2	C3	C4
F1	V1	V2	V3	V4
F2	V4	V3	V2	V1
F3	V2	V4	V1	V3
F4	V3	V1	V4	V2

Solucion:

	C1	C2	C3	C4	Y.j
F1	45	50	43	35	173
F2	29	53	41	63	186
F3	37	41	41	63	182
F4	38	40	35	41	154
Yi.	149	184	160	202	695

V1	V2	V3	V4	
189	169	197	140	695

CALCULO DE SUMAS DE CUADRADOS

Termino de corrección TC = $695^2/16 = 30189.1$

SC(Total) = $45^2 + 50^2 + ... 41^2$ - TC = 1359.9375

SC(Filas) = $(173^2 + ... + 154^2) / 4$ - TC = 152.18750

SC(Columna) = $(149^2 + ... + 202^2) / 4$ – TC = 426.18750

SC(Melon) = $(189^2 + ... + 140^2) / 4$ – TC = 483.68750

SC(error) = SC(total) – SC(filas) – SC(columnas) = 297.8750

Promedio = 695 /16 = 43.438

CM (error) = SC(error) / [(t-1)(t-2)] = 49.6458

CV = Raiz (CM error) *100 / Promedio = 16.2 %

Analisis de Variancia:

Fuente	Gl	S.C.	C.M.	Fc	Pr > F
FILA	3	152.1875	50.7291	1.02	0.4466
COLUMNA	3	426.1875	142.0625	2.86	0.1264
MELON	3	483.6875	161.2291	3.25	0.1022
Error	6	297.8750	49.6458		
Total	15	1359.9375			

Resultado : No existe diferencias estadisticas entre las variedades de melon tratadas con el sistema de riego por exudación.

El coeficiente de variacion es 16 % aceptable para una evaluación en campo.

El rendimiento promedio del melon en condiciones experimentales resulto 43.3 kilos por parcela experimental.

El rendimiento por hibrido fue el siguiente :

V1 : Híbrido Mission 47.3 kilos
V2 : Híbrido Mark. = 42.3 kilos
V3 : Híbrido Topfligth. 49.3 kilos
V4 : Híbrido Hales Best Jumbo. = 35.0 kilos

Según los resultados experimentales no existen diferencias estadisticas entre las variedades ; las diferencias se dan a nivel de 0.10 de probabilidad de error en el rechazo de la hipotesis de nulidad, esto significa que muy posible existen diferencias pero en este experimento no fue posible detectar por los pocos grados de libertad para el error, lo recomendable cuando se prueba un testigo con pocos grados de libertad, se recomienda tener mas repeticiones, en el cuadrado latino, lo recomendable es doblar el numero de parcelas del testigo ; esto significa tener en forma ficticia 5 variedades en 5 filas y 5 columnas y los grados de libertad para el error serian 4x3 = 12. En el analisis de variancia se realiza en forma normal, solo que para las pruebas estadisticas, realizar los contrastes ortogonales apropiados porque el testigo esta representado por dos trataminentos iguales.

Ejemplo : Suponga en este caso que la variedad V4 se dobla en el experimento y se identifica como (V4 y V5), entonces un plan podria ser :

	C1	C2	C3	C4	C5
F1	V1	V2	V3	V4	V5
F2	V4	V3	V2	V5	V1
F3	V2	V4	V5	V1	V3
F4	V3	V5	V1	V2	V4
F5	V5	V1	V4	V3	V2

El Analisis de variancia tendra las siguientes fuentes y grados de libertad :

Fila	4
Columna	4
Melon	4
Error	12
Total	24

La suma de cuadrados de Melon debe ser descompuesta en :

Melon	4
V4, V5 vs V1, V2, V3	1
V4 vs V5	1
Entre V1, V2 , V3	2

Para el ejercicio de este proceso, suponga los siguientes totales de Variedades para el ejemplo

V1	V2	V3	V4	V5	Y(k)
189	169	197	140	145	840

SC(Melón) = $(189^2 + \ldots + 145^2) / 5 - 840^2 / 25 = 519.2$
SC(V4, V5 vs V1, V2, V3) = $(189+169+197)^2 / 15 + (140+145)^2 / 10 - 840^2 / 25 = 433.5$
SC(V4 vs V5) = $140^2 / 5 + 145^2 / 5 - (140 + 145)^2 / 10 = 2.5$
SC(V1, V2 , V3) = $(189^2 + 169^2 + 197^2) / 5 - (189+169+197)^2 / 15 = 83.2$

Este ultimo resultado puede ser obtenido por diferencia del total de SC(Melón)

$519.2 - (433.5 + 2.5) = 83.2$

Para estos calculos tambien puede utilizar los contrastes ortogonales.

	V1	V2	V3	V4	V5	Suma	Numerador	Denominador	SC
V4,V5 vs demas	2	2	2	-3	-3	255	65025	150	433.5
V4 vs V5	0	0	0	1	-1	-5	25	10	2.5
	189	169	197	140	145	840			

Programa R para el analisis del experimento con hibridos de Melón.

```
# cargar las bibliotecas requeridas
library(combinat)
library(agricolae)
#
# Leer los datos del archivo melon.txt
# Factores : fila columna variedad
# variable : rdto
# Diseño cuadrado latino 4x4

datos <- read.table("melon.txt",header=TRUE)
#
attach(datos)
conjunto <-data.frame(fila=factor(fila),columna=factor(columna),variedad=factor(variedad),rdto=rdto)
modelo <-aov(rdto ~ fila + columna + variedad,conjunto)
print(modelo)
print(anova(modelo))

# Utilizar el procedimiento taller

gl<-df.residual(modelo) # extrae los grados del libertad delr error
error<-deviance(modelo)/gl # extrae el CM(error)
compara<-HSD.test(rdto,variedad,gl,error) # realiza las comparaciones multiples de Tukey
```

Análisis del Diseño de parcelas divididas.

Diseño SPLIT PLOT en Bloques

Diseño y datos del campo, Split plot , 6 parcelas grandes distribuidad en 3 bloques, los niveles de A (2) en las parcelas grandes. Cada parcela dividida en 3 subparcelas al cual se aplico el factor B (3 niveles).

Diseño:

		a2				a1	
I	b2	b1	b3		b1	b3	b2

		a1				a2	
II	b2	b3	b1		b2	b1	b3

		a1				a2	
II	b1	b3	b2		b1	b2	b3

Numeracion de las parcelas:

I ->	P1	P2
II ->	P3	P4
III ->	P5	P6

Datos ordenados en el archivo: parcela.txt

Programa R para el analisis estadistico.

```
# cargar las bibliotecas requeridas
library(combinat)
library(agricolae)
#
# Leer los datos del archivo maderable.txt
# factores : bloque parcela A B
# variable : rdto
# Diseño en parcelas divididas
# borrar todos los objetos del espacio de trabajo
rm(list=ls())

datos <-read.table("parcelas.txt",header=TRUE)
attach(datos)
conj6<-data.frame(bloque=factor(bloque),A=factor(A) ,
B=factor(B), parcela<-factor(parcela),rdto=rdto)

modelo <-aov(rdto~bloque+A*B+Error(parcela),data=conj6)
summary(modelo)

# Interprete el analisis de variancia obtenido con la funcion summary()

# extraiga informacion del modelo

gl.error.a<-modelo$parcela$df.residual
```

```
gl.error.b<-modelo$Within$df.residual
x<-modelo$parcela$residuals
y<-modelo$Within$residuals
sc.error.a<-sum(x^2)
sc.error.b<-sum(y^2)
cm.a<-sc.error.a/gl.error.a
cm.b<-sc.error.b/gl.error.b

# Determina el promedio general del experimento
promedio <- mean(conj6$rdto)

# Determina el numero de niveles de los factores
a<-length(levels(conj6$A))
b<-length(levels(conj6$B))

# Halla los coeficientes de variacion
cv.a <- sqrt(cm.a/b)*100/promedio
cv.b <- sqrt(cm.b)*100/promedio

# Realiza las comparaciones de promedios.

compara.a<-LSD.test(conj6$rdto,conj6$A,gl.error.a,cm.a)
compara.b<-LSD.test(conj6$rdto,conj6$B,gl.error.b,cm.b)
```

Análisis del Diseño de bloques dividido.

Diseño STRIP PLOT en Bloques. El diseño consiste en aplicar por filas un tipo de tratamiento y por columnas otra aplicación. El interes principal es la interaccion de estos dos factores (Trat * Aplic), en segundo lugar son los analisis especifico de cada factor en forma independiente.

Para el analisis se tiene 3 errores,

Error para el factor Trat.
Error para el factor Aplic.
Error para la interaccion.

Diseño y datos del campo, Strip plot 4x3 con 2 repeticiones.

Bloque = 1					Bloque = 2				
	Aplic =3	Aplic =1	Aplic =4	Aplic =2		Aplic =2	Aplic =3	Aplic =4	Aplic =1
Trat =1	6.0	9.0	4.0	14.0	Trat =1	15.0	8.0	2.0	8.0
Trat =2	4.0	7.0	6.0	13.0	Trat =3	14.0	7.0	5.0	7.0
Trat =3	5.0	8.0	5.0	12.0	Trat =2	17.0	9.0	3.0	7.0

Para el registro de los datos, se forman 5 columnas:

Bloques, trat, Aplic y rdto.

El archivo: bloques divididos.txt tiene entonces la siguiente informacion:

```
bloque  trat    Aplic      rdto
     1     1     3          6
     1     1     1          9
...
```

son 24 filas.

Programa R para el analisis:

```
# Diseño en parcelas divididas
# Leer los datos del archivo bloques divididos.txt
# factores : bloque  trat    aplic
# variable : rdto

rm(list=ls())

datos <-read.table("bloques divididos.txt",header=TRUE)
attach(datos)
conjunto<-data.frame(bloque=factor(bloque),trat=factor(trat),aplic=factor(aplic),
rdto=rdto)

modelo1 <-aov(rdto~bloque+trat+ Error(bloque:trat),data=conjunto)
modelo2 <-aov(rdto~aplic+ Error(bloque:aplic)+bloque,data=conjunto)
modelo3 <-aov(rdto~bloque:trat + Error(bloque:aplic)+ aplic:trat,data=conjunto)
```

```
> summary(modelo1$"bloque:trat")
          Df Sum Sq Mean Sq F value Pr(>F)
bloque     1  3.375   3.375  3.0000 0.2254
trat       2  0.750   0.375  0.3333 0.7500
Residuals  2  2.250   1.125

> summary(modelo2$"bloque:aplic")
          Df Sum Sq Mean Sq F value  Pr(>F)
aplic      3 330.12  110.04 14.2757 0.02787 *
bloque     1   3.37    3.37  0.4378 0.55545
Residuals  3  23.12    7.71
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

> summary(modelo3$Within)
            Df  Sum Sq Mean Sq F value Pr(>F)
bloque:trat  4  3.0000  0.7500  0.7826 0.5757
trat:aplic   6 11.2500  1.8750  1.9565 0.2172
Residuals    6  5.7500  0.9583
>

# Comparacion de tratamientos
> library(combinat)
> library(agricolae)
> gl<-df.residual(modelo1$"bloque:trat")
> error<-deviance(modelo1$"bloque:trat")/gl
> compara<-LSD.test(rdto,trat,gl,error)

# Comparacion de Aplicaciones
> gl<-de.residual(modelo2$"bloque:aplic")
> error<-deviance(modelo2$"bloque:aplic")/gl
> compara<-LSD.test(rdto,aplic,gl,error)
```

```
# Promedio y coeficiente de variacion

> promedio<-mean(rdto)
> gl <-df.residual(modelo3$Within)
> error <-deviance(modelo3$Within)/gl
> cv<- sqrt(error)*100/promedio
> cv
[1] 12.04855
> promedio
[1] 8.125
```

Experimentos repetidos en espacio y tiempo.

```
Cultivo de algodon 4 linajes.

Diseño bloques completos al azar (6 bloques).

Cultivo:
   1 = Plantada (1er Año)
   2 = Soca     (2do Año)

Lugares
   1 = Valle de Lima
   2 = Valle de Pisco
```

Datos.

Valle de Lima							
Linaje	Cultivo	I	II	III	IV	V	VI
1	1	12	13	11	13	11	15
	2	15	20	17	18	19	23
2	1	10	12	15	11	13	13
	2	14	16	17	14	17	26
3	1	9	9	12	11	13	12
	2	13	11	17	12	14	13
4	1	8	12	4	8	12	9
	2	9	16	7	11	14	12

Valle de Pisco							
Linaje	Años	I	II	III	IV	V	VI
1	1	9	18	11	20	14	17
	2	10	22	20	27	21	21
2	1	19	13	14	26	13	17
	2	17	24	21	23	21	21
3	1	12	16	14	17	22	14
	2	22	19	22	22	22	20
4	1	15	20	15	12	14	12
	2	14	24	20	24	19	19

Programa R para el analisis completo

```
# cargar las bibliotecas requeridas
#
# analisis en diferentes lugares y epocas.
# en el mismo campo en ambas epocas, es decir dos cosechas del mismo cultivo
# Leer los datos del archivo bloques divididos.txt
# factores : lugar   bloque  Linaje  epoca
# variable : rdto

rm(list=ls())

datos <-read.table("algodon.txt",header=TRUE)
attach(datos)
conjunto<-data.frame(lugar=factor(lugar),bloque=factor(bloque),epoca=factor(epoca),
linaje=factor(linaje),rdto=rdto)

# Seleccion de la informacion.
attach(conjunto)
epoca.1 <-subset(conjunto,epoca==1)
epoca.2 <-subset(conjunto,epoca==2)
sitio.lima <-subset(conjunto,lugar=="Lima")
sitio.pisco <-subset(conjunto,lugar=="Pisco")
attach(sitio.lima)
sitio.lima.1 <- subset(sitio.lima, epoca==1)
sitio.lima.2 <- subset(sitio.lima, epoca==2)
attach(sitio.pisco)
sitio.pisco.1 <- subset(sitio.pisco, epoca==1)
sitio.pisco.2 <- subset(sitio.pisco, epoca==2)
```

analisis en la primera y segunda epoca, en lima.

```
modelo1.1 <-aov(rdto~bloque+linaje,data=sitio.lima.1)
modelo1.2 <-aov(rdto~bloque+linaje,data=sitio.lima.2)
```

analisis combiando por epocas en lima

```
modelo1 <- aov(rdto~epoca + bloque  + Error(epoca:bloque)+linaje + linaje:epoca, data=sitio.lima)
```

analisis en la primera y segunda epoca, en pisco.

```
modelo2.1 <-aov(rdto~bloque+linaje,data=sitio.pisco.1)
modelo2.2 <-aov(rdto~bloque+linaje,data=sitio.pisco.2)
```

analisis combiando por epocas en pisco

```
modelo2 <- aov(rdto~epoca + bloque + Error(epoca:bloque)+linaje + linaje:epoca,
data=sitio.pisco)
```

analisis combinado por lugares en la primera epoca.

```
modelo3.1 <-aov(rdto~lugar+Error(bloque%in%lugar) +linaje +linaje:lugar,data=epoca.1)
# analisis combinado por lugares en la segunda epoca.

modelo3.2 <-aov(rdto~lugar+Error(bloque%in%lugar) +linaje +linaje:lugar,data=epoca.2)
```

analisis combinado por epocas y lugares

```
modelo4.1 <-aov(rdto~lugar+Error(bloque%in%lugar),data=conjunto)
modelo4.2 <-aov(rdto~lugar+bloque%in%lugar+linaje+lugar:linaje + Error(linaje:bloque%in%lugar)
+ epoca+lugar:epoca + epoca:bloque%in%lugar +linaje:epoca + lugar:linaje:epoca ,data=conjunto)
```

Resultados:

analisis en la primera y segunda epoca, en lima.

```
> summary(modelo1.1)
            Df Sum Sq Mean Sq F value  Pr(>F)
bloque       5 20.333   4.067  0.9337 0.48697
linaje       3 51.667  17.222  3.9541 0.02916 *
Residuals   15 65.333   4.356
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
> summary(modelo1.2)
            Df  Sum Sq Mean Sq F value   Pr(>F)
bloque       5  81.708  16.342  1.8886 0.156167
linaje       3 202.458  67.486  7.7994 0.002276 **
Residuals   15 129.792   8.653
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
>
```

analisis combiando por epocas en lima

```
> summary(modelo1)

Error: epoca:bloque
          Df  Sum Sq Mean Sq F value    Pr(>F)
epoca      1 196.021 196.021 68.2801 0.0004234 ***
bloque     5  87.687  17.537  6.1089 0.0344110 *
Residuals  5  14.354   2.871
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Error: Within
             Df  Sum Sq Mean Sq F value    Pr(>F)
linaje        3 223.396  74.465 11.4489 3.591e-05 ***
epoca:linaje  3  30.729  10.243  1.5748     0.216
Residuals    30 195.125   6.504
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
>
```

analisis combinado por epocas y lugares

```
> summary(modelo4.1$"bloque:lugar")
          Df Sum Sq Mean Sq F value   Pr(>F)
lugar      1 580.17  580.17  20.107 0.001172 **
Residuals 10 288.54   28.85
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
> summary(modelo4.2$"linaje:bloque:lugar")
```

```
Del analisis retirar las fuentes:
Lugar y Lugar:bloque que es el error para el analisis del primer ANVA

             Df Sum Sq Mean Sq F value   Pr(>F)
lugar         1 580.17  580.17 48.8962 9.05e-08 ***
linaje        3 149.21   49.74  4.1917  0.01365 *
lugar:bloque 10 288.54   28.85  2.4318  0.02928 *
lugar:linaje  3  99.08   33.03  2.7836  0.05792 .
Residuals    30 355.96   11.87
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
> summary(modelo4.2$"Within")
                   Df Sum Sq Mean Sq F value    Pr(>F)
epoca               1 495.04  495.04 76.9661 8.795e-10 ***
lugar:epoca         1   6.00    6.00  0.9328    0.3418
linaje:epoca        3  13.54    4.51  0.7018    0.5584
lugar:epoca:bloque 10  43.21    4.32  0.6718    0.7411
lugar:linaje:epoca  3  20.25    6.75  1.0494    0.3851
Residuals          30 192.96    6.43
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
>
```

Aplicación en campo de agricultores.

Considere los siguientes resultados de un experimento con diferentes agricultores en Colombia, en el Oriente antioqueño (1972) (ICA.- Aplicación por Orlando Martinez W.)

Rendimiento de Frijol en kg/ha. En grano limpio y seco.
Tecnologías:
20-40-20 kg/ha. de N. P_2O_5 y K_2O + 2 t/ha de gallinaza.
40-80-40 kg/ha. de N. P_2O_5 y K_2O + 2 t/ha de gallinaza.
60-120-60 kg/ha. de N. P_2O_5 y K_2O + 2 t/ha de gallinaza.
40-80-40 kg/ha. de N. P_2O_5 y K_2O + 4 t/ha de gallinaza.

Resultados:

Municipio	Parcela	a	b	c	d	Promedio
Rionegro	Capiro	1036	1272	787	806	975
Rionegro	Cristo_Rey	364	249	258	370	310
Carmen	Garzonas	1019	1019	911	1215	1041
Carmen	Palmas	447	780	717	331	569
Carmen	Palmas	1532	1214	1246	1530	1381
Carmen	Campo_alegre	1161	1521	1410	1379	1368
Carmen	Chapa	1309	1477	1626	1657	1517
Carmen	Chapa	834	1087	1065	971	989
Carmen	Chapa	729	787	844	806	792
Carmen	Soñadora	749	543	700	672	666
Carmen	Soñadora	918	1214	1055	929	1029
Carmen	Quiramana	668	835	818	791	778
Carmen	Quiramana	460	566	508	470	501
La_Union	La_Maria	1446	1794	1688	1731	1665

Marinilla	Asuncion	770	527	960	506	691
Marinilla	Asuncion	675	992	401	728	699
Marinilla	Llanadas	515	606	422	448	498
Marinilla	Llanadas	2110	1987	1706	2137	1985
Marinilla		1046	1993	958	1090	1272
Marinilla		517	559	527	612	554
Gurane	San_Jose	1002	1302	1020	985	1077
	Promedio	919	1063	935	960	969

Regresion	Intercepto	5.43	46.57	57.13	-109.13
	Pendiente	0.94	1.05	0.91	1.10
	R2	0.93	0.84	0.88	0.95
	CM regresión	3356725	4152047	3094180	4594967

En la prueba estadística de los coeficientes de regresión, ninguno rechaza la hipótesis nula, sin embargo, los mas cercanos a la unidad es el tratamiento (a y b), respecto al CM de la regresión, los mas bajos son los tratamientos (a, b y c), d es muy alto; y respecto al promedio, el tratamiento con promedio alto es (b)

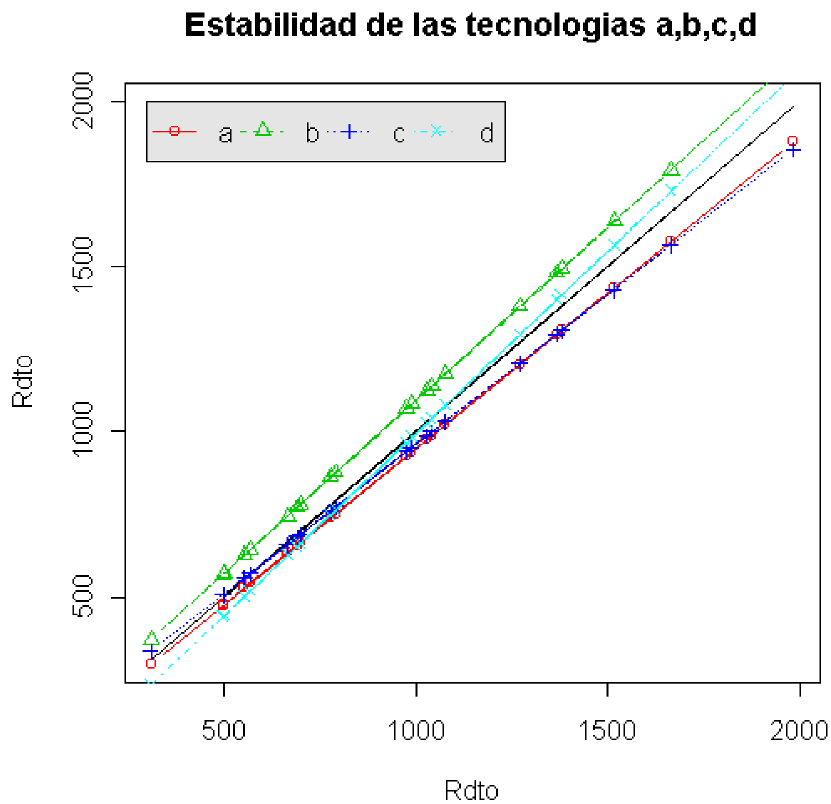
Programa R para el estudio de las tecnologias:

```
# cargar las bibliotecas requeridas
#
# Leer los datos del archivo agricultor.txt
# factores : municipio y parcela
# variable : rendimiento
# Evaluacion de 4 tecnologias en campo de agricultores

datos <- read.table("agricultor.txt",header=TRUE)
#
attach(datos)
conjunto <-data.frame(a=a,b=b,c=c,d=d)
matriz <- as.matrix(conjunto)

# se calcula el indice ambienta como el promedio de las tecnologias.

promedio <- apply(matriz,1,mean)
matriz <-cbind(matriz,promedio)

# Determina los modelos de regresion lineal simple por tecnologia
lm.a<-lm(matriz[,1]~matriz[,5])
lm.b<-lm(matriz[,2]~matriz[,5])
lm.c<-lm(matriz[,3]~matriz[,5])
lm.d<-lm(matriz[,4]~matriz[,5])

# como obtener R² del modelo
summary(lm.a)$"r.squared"

# calcula las desviaciones estandar de los coeficientes
n<-nrow(matriz)
sx<-(n-1)*var(matriz[,5])
```

```
error <-deviance(lm.a)/df.residual(lm.a)
sb.a<-sqrt(error/sx)
error <- deviance(lm.b)/df.residual(lm.b)
sb.b<-sqrt(error/sx)
error <- deviance(lm.c)/df.residual(lm.c)
sb.c<-sqrt(error/sx)
error <- deviance(lm.c)/df.residual(lm.c)
sb.d<-sqrt(error/sx)

# extrae los coeficientes beta de los modelos
b.a<-as.vector(lm.a$coefficients)[2]
b.b<-as.vector(lm.b$coefficients)[2]
b.c<-as.vector(lm.c$coefficients)[2]
b.d<-as.vector(lm.d$coefficients)[2]

# Realiza las hipotesis sobre la pendiente, es TRUE si se acepta la
hipotesis que b=1
ta<-abs(b.a - 1)/sb.a
tb<-abs(b.b - 1)/sb.b
tc<-abs(b.c - 1)/sb.c
td<-abs(b.d - 1)/sb.d

t0.05<-qt(0.975,n-2)

cat("Tecnologia 'a': " , ta < t0.05,"\n")
cat("Tecnologia 'b': " , tb < t0.05,"\n")
cat("Tecnologia 'c': " , tc < t0.05,"\n")
cat("Tecnologia 'd': " , td < t0.05,"\n")
#---------------termina este proceso ---------------
```

Se puede afirmar que el tratamiento b cumple con la condición de tratamiento estable y tiene mejores respuestas.

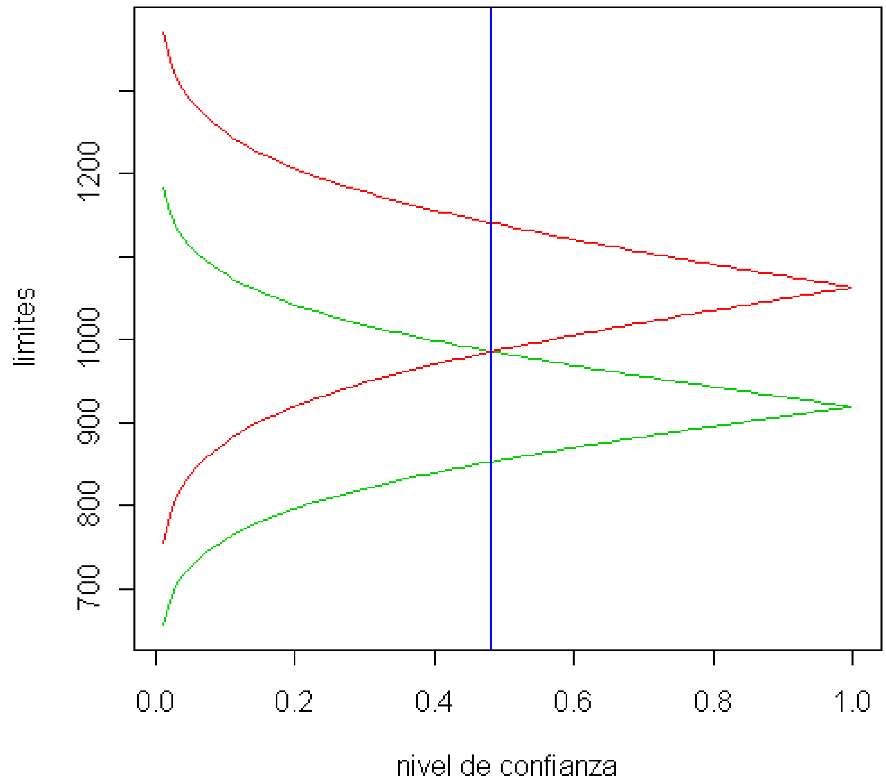
El gráfico de las líneas de regresión



Procedimiento R para comparar graficamente las pendientes.

```
# Grafica las pendientes junto a la perfecta estabilidad
plot(x,x,cex=0.4,type="l",xlab="Rdto",ylab="Rdto",
 main=("Estabilidad de las tecnologias a,b,c,d"),lty=1)
lines(x,y.a,col=2,pch = 1,lty=c(1),type="b")
lines(x,y.b,col=3,pch = 2,lty=c(2),type="b")
lines(x,y.c,col=4,pch = 3,lty=c(3),type="b")
lines(x,y.d,col=5,pch = 4,lty=c(4),type="b")
max.y <- 2000
min.x <- 300
legend(min.x,max.y, c("a", "b","c", "d"), col = c(2,3,4,5),
lty = c(1,2,3,4), pch = c( 1, 2,3,4), merge = TRUE, bg='gray90',horiz=TRUE)
```

Para comprobar si el tratamiento b difiere del tratamiento a , se debe hallar los intervalos de confianza.



Procedimiento R para comparar 2 tecnologias mediante limites de confianza

```
# comparacion de dos tecnologias mediante los limites de confianza
# Limites de confianza de 2 tecnologias; por ejemplo "a" vs "b"
promedio.a <-mean(matriz[,1])
var.a<-var(matriz[,1])
promedio.b =mean(matriz[,2])
var.b<-var(matriz[,2])
k<-1
y<-rep(100,0); li.a<-y ; ls.a<-y; li.b<-y ; ls.b<-y
for (i in 1:100) {
k<-k-0.01
y[i]<-1-k; x<-1-(1-k)/2
li.a[i]<-promedio.a -qt(x,n-1)*sqrt(var.a/n)
ls.a[i]<-promedio.a +qt(x,n-1)*sqrt(var.a/n)
li.b[i]<-promedio.b -qt(x,n-1)*sqrt(var.b/n)
ls.b[i]<-promedio.b +qt(x,n-1)*sqrt(var.b/n)
}
limites<-cbind(li.a,ls.a,li.b,ls.b)
matplot(y,limites,col=c(3,3,2,2),type=c("l","l","l","l"),
xlab="nivel de confianza",lty=c(1,1,1,1))
abline(v=0.48,col="blue")
```

El área de los limites de confianza de los tratamiento A y B se superponen hasta cerca 0.45, significa que el riesgo del rechazo es mayor de 0.05 significa que no existe diferencia entre las tecnologias “a” y “b”

Referencia bibliografica

1. Calzada B. José (1970) Métodos estadísticos para la investigación. Editorial Juridica S.A. Lima, Perú.
2. Ching Chun Li (1969) Introducción a la estadística experimental. Ediciones Omega, S. A. Barcelona.
3. Cochran, William & Cox, Gertrude (1950) Experimental Design, John Wiley & Sons. INC. New York.
4. Ihata R. & Gentleman R. 1996. R: a languaje for data analysis and graphics. Journal of Computational and Graphical Statistics 5: 299-314.
5. Kwanchai A. Gomez (1985) Statistical procedures for agricultural research. Second edition. John Wiley & Sons. INC. New York.
6. LeClerg, Erwin; Warren, Leornad Field plot terchnique. Second Edition. Burgess Publishing Company. Minneapolis 23, Minnesota.
7. Ostle, Bernard (1992) Estadistica Aplicada. Editorial Limusa, S.A. México D.F.
8. Steel, Robert & Torrie, James & Dickey, David “Principles and Proceduires of Statistics a Biometrical Approach” Third Edition 1997. Mc Graw Hill Series In Probability and Statistics