

UNIVERSIDAD DE LA FRONTERA
FACULTAD DE CIENCIAS AGROPECUARIAS Y FORESTALES



**COMPARACIÓN DE LA BONDAD DE AJUSTE DE MODELOS MIXTOS NO
BALANCEADOS MEDIANTE TRES SOFTWARE DE ANÁLISIS ESTADÍSTICO.**

Tesis presentada a la Facultad de Ciencias Agropecuarias y Forestales de la Universidad de La Frontera. Como parte de los requisitos para optar al título de Ingeniero Agrónomo

GREGORI HERNAN EHIJOS ARAVENA

TEMUCO – CHILE

2013

UNIVERSIDAD DE LA FRONTERA
FACULTAD DE CIENCIAS AGROPECUARIAS Y FORESTALES



**COMPARACIÓN DE LA BONDAD DE AJUSTE DE MODELOS MIXTOS NO
BALANCEADOS MEDIANTE TRES SOFTWARE DE ANÁLISIS ESTADÍSTICO.**

Tesis presentada a la Facultad de Ciencias Agropecuarias y Forestales de la Universidad de La Frontera. Como parte de los requisitos para optar al título de Ingeniero Agrónomo

GREGORI HERNAN EHIJOS ARAVENA

PROFESOR GUÍA: HORACIO JULIO MIRANDA VARGAS

TEMUCO – CHILE

2013

AGRADECIMIENTOS.

Es difícil resumir en pocas palabras lo que ha ocurrido durante estos años de estudio, y poder agradecer con justicia todo el apoyo y cariño que he recibido, en especial en estos últimos tres años, que han sido particularmente difíciles, pero que al mismo tiempo han servido para dar sentido y propósito al esfuerzo realizado.

Quisiera agradecer en primer lugar a mi madre, porque aún en momentos muy dolorosos siguió apoyándome incondicionalmente, y porque en muchas ocasiones, su recuerdo fue la motivación que me hizo seguir a pesar de las dificultades. Sin ella no habría podido lograr esta meta.

A mi hermana, porque siempre tuvo fe en mí, y porque me apoyó y acompañó en los momentos difíciles, también a mi padre, que desde niños despertó en nosotros la curiosidad por aprender más allá del contexto de una sala de clases, y a Marilyn, que con su serenidad y ternura me contuvo en momentos de aflicción y soledad.

A los profesores Horacio Miranda y Christian Salas, por su paciencia y comprensión durante el desarrollo de esta trabajo, y por el privilegio de haber realizado mi trabajo de tesis con ellos, al profesor Jorge Barahona, por su buena disposición y empatía, que reflejan una gran calidad humana, y a la profesora Sonia Salvo, del Departamento de Matemática y Estadística, por su ayuda y valiosos consejos el día del examen.

Y a todos quienes lo breve de este texto no permite incluir, pues durante este tiempo he sentido el cariño y apoyo desinteresado de muchas personas, lo que me hace sentir muy agradecido de la vida que me ha tocado vivir.

ÍNDICE

Capítulo		Página
1	INTRODUCCIÓN.	1
2	REVISIÓN BIBLIOGRÁFICA.	2
2.1	La investigación experimental en ciencias biológicas.	2
2.1.1	Diseños experimentales comúnmente empleados en investigación biológica.	3
2.2	Modelación de los efectos.	3
2.2.1	Modelos de efectos fijos.	4
2.2.2	Modelos de efectos aleatorios.	5
2.2.3	Modelos mixtos.	6
2.2.4	Modelos balanceados y no balanceados.	7
2.2.5	Factores de corrección.	8
2.3	Estimación de parámetros.	8
2.3.1	Estimaciones de μ y β .	9
2.3.2	Mínimos cuadrados ponderados.	10
2.3.3	Máxima verosimilitud.	10
2.4	Estimación de componentes de varianza.	11
2.4.1	Máxima verosimilitud restringida.	11
2.4.2	MINQUE y MIVQUEO.	11
2.5	Predicción de factores aleatorios.	12
2.6	Estimación de sumas de cuadrados.	12
2.6.1	Sumas de cuadrados tipo I.	13
2.6.2	Sumas de cuadrados tipo II.	14
2.6.3	Sumas de cuadrados tipo III.	15
2.6.4	Sumas de cuadrados tipo VI.	16
2.7	Evaluación del ajuste de los modelos.	16
2.7.1	Criterio de Información de Akaike (AIC).	17
	Criterio de Información de Akaike Corregido (AIC _C).	18
2.7.2	Criterio de Información Bayesiano (BIC).	19
3	MATERIALES Y MÉTODOS.	20

3.1	Materiales.	20
3.1.1	Diseño de un factor con bloque y covarianza.	20
3.1.2	Diseño de un factor fijo y un factor aleatorio.	22
3.1.3	Diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.	24
3.1.4	Diseño de parcelas divididas.	26
3.1.5	Diseño cross – over.	28
3.2	Métodos.	30
4	PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS.	31
4.1	Diseño de un factor con bloque y covarianza.	32
4.1.1	Resultados de los análisis.	32
4.2	Diseño de un factor fijo y un factor aleatorio.	33
4.2.1	Resultados de los análisis.	34
4.3	Diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.	35
4.3.1	Resultados de los análisis.	35
4.4	Diseño de parcelas divididas.	37
4.4.1	Resultados de los análisis.	37
4.5	Diseño cross – over.	39
4.5.1	Resultados de los análisis.	39
5	CONCLUSIONES.	41
6	RESUMEN.	42
7	SUMMARY.	43
8	LITERATURA CITADA.	44
9	ANEXOS	54

ÍNDICE DE CUADROS

Cuadro		Página
1	Tabla de ANOVA teórica bloques al azar.	21
2	Estructura de la tabla de datos para el ejemplo de un factor con bloque y covarianza.	22
3	Tabla de ANOVA teórica dos factores Completamente Aleatorizado.	23
4	Estructura de la tabla de datos para el ejemplo de un factor fijo y un factor aleatorio.	24
5	Tabla de ANOVA teórica Parcelas Dividida en Bloque completos al azar.	24
6	Estructura de la tabla de datos para el ejemplo de un factor fijo y un factor aleatorio.	25
7	Tabla de ANOVA teórica Parcelas Dividida en Bloque completos al azar.	26
8	Estructura de la tabla de datos para el ejemplo de parcelas divididas.	27
9	Tabla de ANOVA teórica parcelas dividida completamente al azar.	28
10	Estructura de la tabla de datos para el ejemplo de diseño cross – over.	29
11	Estimaciones de componentes de varianza.	32
12	Valores F y probabilidades.	32
13	Evaluación del ajuste de los modelos.	33
14	Estimaciones de componentes de varianza.	34
15	Valores F y probabilidades.	34
16	Evaluación del ajuste de los modelos.	35
17	Estimaciones de componentes de varianza.	35
18	Valores F y probabilidades.	36
19	Evaluación del ajuste de los modelos.	36
20	Estimaciones de componentes de varianza.	37
21	Valores F y probabilidades.	38

22	Evaluación del ajuste de los modelos.	38
23	Estimaciones de componentes de varianza.	39
24	Valores F y probabilidades.	39
25	Evaluación del ajuste de los modelos.	40
26	Cuadro comparativo de los criterios de información de Akaike y Bayesiano en SAS, R y JMP.	40

ÍNDICE DE FORMULAS

Formula		Página
1	Modelo de efectos fijos.	4
2	Modelo de efectos aleatorios.	6
3	Modelo de efectos mixtos.	7
4	Ecuaciones de mínimos cuadrados ordinarios.	9
5	Criterio de Información de Akaike.	18
6	Criterio de Información de Akaike Corregido.	18
7	Criterio de información Bayesiano.	19

1. INTRODUCCIÓN.

La estadística tiene aplicaciones en casi todas las áreas del saber, y permite a los investigadores contar con una herramienta que les hace posible estudiar, con alto grado de certeza, fenómenos en los que la multiplicidad de variables que inciden dificultan la aplicación de otro tipo de metodología. La agricultura no es ajena al uso de la estadística y se ve muy favorecida por sus métodos, debido a la intrínseca variabilidad que presentan las entidades biológicas y la compleja y multifactorial interacción que tienen con el medio que las rodea, la que puede ser adecuadamente representada mediante modelos estadísticos.

Es en este contexto que las primeras aplicaciones de la estadística a las ciencias biológicas comenzaron en las ciencias agropecuarias, dando origen a la bioestadística.

En la investigación agrícola se analizan los diseños experimentales mediante modelos lineales, conteniendo estos generalmente al menos un factor aleatorio, además del error experimental, y teniéndose en muchos casos conjuntos de datos no balanceados.

Por lo tanto, el presente trabajo de investigación tiene como propósito realizar las estimaciones de diferentes modelos de análisis de varianza habitualmente empleados en investigación agropecuaria, con la finalidad de evaluar el grado de similitud obtenido por medio de diferentes programas estadísticos empleados en investigación.

Para realizar lo anterior, se recurrió a un conjunto de diseños experimentales de uso común en ciencias agropecuarias y se seleccionaron ejemplos que estuvieran disponibles en Internet, basados en resultados obtenidos mediante su análisis con el software estadístico de SAS Institute Inc., que fue usado como el programa computacional de referencia para realizar las comparaciones.

2. REVISIÓN BIBLIGRÁFICA.

2.1. La investigación experimental en ciencias biológicas.

En ciencias biológicas tradicionalmente se realiza investigación de tipo empírica, en la cual se utiliza el razonamiento inductivo, que permite, mediante el estudio de muestras de una población, llegar a conclusiones que se pueden generalizar a todos los individuos de dicha población (Miranda, 2011a). Para la realización de la investigación empírica anteriormente descrita se suele recurrir a diseños experimentales, que permiten la comparación de promedios entre grupos o tratamientos pertenecientes a uno o más factores controlados (factores fijos). Tales diseños pueden clasificarse según diversos criterios, como; el número de factores involucrados en el experimento, las posibilidades de aleatorización de los niveles de aquellos factores (tratamientos), los distintos tipos de interacción posible entre dos o más factores, y la adición de errores experimentales especiales en el diseño.

En relación al número de factores involucrados, estos pueden considerar diseños de un factor, en que hay una sola variable categórica que determina los grupos a comparar, o diseños factoriales, que poseen dos o más variables categóricas (Quinn y Keough, 2002).

En cuanto a la aleatorización, se pueden utilizar diseños experimentales completamente aleatorizados, donde los niveles de él o los factores involucrados se asignan en forma totalmente aleatoria a las unidades experimentales, o también se pueden realizar diseños en que se restringe la aleatorización, como el caso en que se utilizan bloques, o en algunos diseños experimentales especiales, como los de parcelas divididas o los de bloques divididos o franjas (Little y Hills, 2002; Walpole *et al.*, 2007).

En lo que respecta a las interacciones, estas se producen en experimentos factoriales, y pueden detectarse cuando hay más de una observación por cada combinación de tratamientos (Spiegel *et al.*, 2010).

Respecto al último criterio, este hace mención a que en determinados diseños surgen errores experimentales adicionales, como en el caso de los diseños de parcelas divididas, en que se tienen dos errores experimentales, uno de ellos asociado al factor principal y el otro asociado al factor secundario (Piepho *et al.*, 2003; Miranda, 2011b).

2.1.1. Diseños experimentales comúnmente empleados en investigación biológica.

Entre los diseños experimentales de uso más frecuente en investigación biológica se encuentran los diseños de uno o más factores, ya sean completamente al azar o con factor de bloqueo (de naturaleza cuantitativa), y adicionalmente al factor de bloqueo se puede tener una o más covariables (de naturaleza cuantitativa continua) (Kuehl, 2001).

Los diseños experimentales más habituales en investigación biológica son; diseños de parcelas divididas, de parcelas subdivididas, de bloques divididos y de cuadro latino, ya sean con arreglo completamente aleatorizado o en bloques completos al azar (Little y Hills, 2002; Rodríguez, 2000).

2.2. Modelación de los efectos.

Los modelos que se describen a continuación, representan la relación entre una o más variables independientes (factores) con una variable dependiente (variable de respuesta). Se trata de modelos estadísticos probabilísticos, que toman en cuenta el comportamiento aleatorio de la variable y mediante la incorporación de un término de error aleatorio ($\varepsilon = \mathbf{y} - E(\mathbf{y})$), proporcionando así una descripción más realista que la obtenida mediante el uso de un modelo matemático determinístico (Wackerly, 2002).

2.2.1. Modelos de efectos fijos. Un efecto fijo corresponde a aquel en donde los niveles estudiados representan todos los posibles niveles del factor, o al menos todos los niveles acerca de los cuales se hace la inferencia (Littell *et al.*, 1996). En este caso los niveles del factor son medidos sin error y representan los únicos niveles de interés en la investigación (Hamer y Simpson, 2002; Dickey, 2008).

Se dice que los efectos de los grupos son fijos porque son específicamente elegidos o definidos mediante un proceso que no es aleatorio (Kaps y Lamberson, 2009), siendo el investigador quien selecciona los niveles de el o los factores involucrados (Morales, 2005).

La representación del modelo de efectos fijos, también conocido como modelo de ANOVA tipo I (Logan, 2010), usando la notación matricial, se presenta a continuación.

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

Formula 1. Modelo de efectos fijos.

La ecuación anterior representa al modelo lineal general (Searle, 1987), donde \mathbf{Y} es el vector de observaciones, \mathbf{X} es la matriz del diseño que relaciona β con \mathbf{y} , β es el vector de parámetros y ε es el vector de errores aleatorios, con media $E(\varepsilon) = \mathbf{0}$, y varianza $Var(\mathbf{y}) = \sigma^2\mathbf{I}$, siendo $\mathbf{0}$ un vector con todos sus elementos igual a cero e \mathbf{I} una matriz identidad. El valor esperado y varianza de \mathbf{y} son: $E(\mathbf{y}) = \mathbf{X}\beta$ y $Var(\mathbf{y}) = \sigma^2\mathbf{I}$ (Kaps y Lamberson, 2009).

2.2.2. Modelos de efectos aleatorios. Se dice que los factores son aleatorios cuando estos son usados en un estudio para representar sólo una muestra (idealmente una muestra aleatoria) de un gran conjunto de potenciales niveles (grupos o tratamientos), los que constituyen una población con distribución de probabilidad (Littell *et al.*, 1996; Kuehl, 2001; Ramroop, 2002; Piepho *et al.*, 2003; Cayuela, 2010).

Para establecer diferencias entre los factores fijos y aleatorios, podríamos señalar que en un factor fijo se tiene un número pequeño de grupos o tratamientos, cada uno de ellos representando una única población con su propia media y donde la variabilidad entre grupos no es explicada por alguna distribución de probabilidad, mientras que en un factor aleatorio se tiene un gran (incluso infinito) número de grupos o tratamientos, donde los grupos investigados son una muestra aleatoria tomada de una única población de grupos, siendo cada grupo particular una variable aleatoria con alguna distribución de probabilidad (Kaps y Lamberson, 2009).

Adicionalmente, mientras las conclusiones de un experimento con efectos fijos se restringen específicamente a los niveles de tratamientos escogidos para el estudio, en un experimento con efectos aleatorios las conclusiones son extrapoladas a todos los posibles niveles de tratamientos (Logan, 2010).

Finalmente, se debe tener en cuenta que los objetivos del análisis de un factor fijo y un factor aleatorio son distintos, ya que en el caso de un factor fijo el objetivo es la estimación de las medias de grupos o tratamientos para su posterior comparación, mientras que en los factores aleatorios el objetivo es la estimación de los componentes de varianza (Dickey, 2008), donde la varianza del efecto aleatorio y la varianza del error residual corresponden a los componentes de varianza entre y dentro de grupos respectivamente (Kuehl, 2001; Kaps y Lamberson, 2009).

En cierto sentido, se podría decir que cada modelo contiene al menos un efecto aleatorio, el cual corresponde al efecto que constituye el error residual (SAS Institute Inc., 2010a).

La representación del modelo de efectos aleatorios para un ANOVA, también conocido como modelo de ANOVA tipo II (Sanni y Ukaegbu, 2012), usando la notación matricial, se presenta a continuación.

$$\mathbf{Y} = \mu + \mathbf{Z}\mathbf{u} + \varepsilon$$

Formula 2. Modelo de efectos aleatorios.

Donde \mathbf{Y} es el vector de observaciones, μ es la media, \mathbf{Z} es la matriz del diseño que relaciona \mathbf{u} con \mathbf{Y} , \mathbf{u} es el vector de efectos aleatorios con media $\mathbf{0}$ y varianza $\mathbf{G} = \sigma_{\tau}^2 \mathbf{I}_a$, y ε es el vector de errores aleatorios, con media $\mathbf{0}$ y varianza $\mathbf{R} = \sigma^2 \mathbf{I}_{an}$, donde σ_{τ} y σ^2 son los componentes de varianza entre y dentro de grupos respectivamente. Se asume que \mathbf{Y} , \mathbf{u} y ε tienen distribución normal y que \mathbf{u} y ε son independientes (Kaps y Lamberson, 2009). Los valores esperados y (co)varianzas de las variables aleatorias son: $E(\mathbf{u}) = \mathbf{0}$, $Var(\mathbf{u}) = \mathbf{G}$, $E(\varepsilon) = \mathbf{0}$, $Var(\varepsilon) = \mathbf{R}$, $E(\mathbf{y}) = \mathbf{0}$, $Var(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$ (Littell *et al.*, 1996).

2.2.3. Modelos mixtos. Los modelos mixtos son aquellos que contienen tanto efectos fijos como aleatorios (Littell *et al.*, 1996; Bates *et al.*, 2012), en los cuáles los efectos fijos explican la media y los efectos aleatorios explican la estructura de varianza – covarianza de la variable dependiente (Kaps y Lamberson, 2009). Se les suele denominar modelo III, y su naturaleza mixta produce que

las interacciones entre efectos fijos y aleatorios sean aleatorias (Walpole *et al.*, 2007). La representación del modelo mixto, usando una notación matricial, se muestra a continuación.

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$$

Formula 3. Modelo de efectos mixtos.

Donde \mathbf{Y} es el vector de observaciones, \mathbf{X} es la matriz del diseño que relaciona β con \mathbf{y} , β es el vector de efectos fijos, \mathbf{Z} es la matriz del diseño que relaciona \mathbf{u} con \mathbf{y} , \mathbf{u} es el vector de efectos aleatorios, con media cero y matriz de varianza – covarianza \mathbf{G} , ε es el vector de errores aleatorios con media cero y matriz de varianza – covarianza \mathbf{R} . Se asume que \mathbf{Y} , \mathbf{u} y ε tienen distribución normal y que \mathbf{u} y ε son independientes, lo que se representa de la siguiente manera: $\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$, $\mathbf{u} \sim N(0, \mathbf{G})$, $\varepsilon \sim N(0, \mathbf{R})$ y $Cov(\mathbf{u}, \varepsilon) = 0$ (Ramroop, 2002; Kaps y Lamberson, 2009). Los valores esperados y varianzas para el modelo mixto son: $E(\mathbf{Y}) = \mathbf{X}\beta$, $Var(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$, $E(\mathbf{u}) = 0$, $Var(\mathbf{u}) = \sigma_u^2 \mathbf{I}_a$, $E(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2 \mathbf{I}_N$ (Cadena y Castillo, 2000).

2.2.4. Modelos balanceados y no balanceados. El término “no balanceado” hace referencia a experimentos en los cuales hay distinto número de observaciones por nivel o combinación de niveles de variables independientes (Kaps y Lamberson, 2009; Spiegel *et al.*, 2010).

También se puede señalar que, si un experimento tiene el mismo número de unidades experimentales asignadas a cada celda, el experimento es balanceado; de otra manera, el diseño es no balanceado (Shaw y Mitchell – Olds, 1993; Macnaughton, 1998; Sanni y Ukaegbu, 2012).

Los diseños balanceados se pueden transformar en no balanceados cuando una o más observaciones se pierden debido a circunstancias imprevistas (Montgomery, 1991; Rodríguez, 2000; Spilke et al., 2005; Kaps y Lamberson, 2009). También pueden surgir por eliminación debido a una respuesta fuera de rango, denominados datos outlier (Miranda, 2011b).

2.2.5. Factores de corrección. En un diseño experimental, las mediciones de la variable de respuesta pueden estar influidas por las condiciones en que se realiza el experimento o por propiedades intrínsecas del material experimental, las que se denominan factores de corrección, y que dificultan una determinación correcta del efecto de él o los factores investigados. Si tales factores de corrección son de tipo cualitativo se les denomina factores de bloqueo, y si son de tipo cuantitativo se les denomina covariables (Miranda, 2011b).

El factor de bloqueo y las covariables se usan en investigación para corregir la variación entre unidades experimentales. El uso de factores de bloqueo consiste en que las unidades experimentales más similares entre sí se agrupan dentro de bloques, donde luego aquellas unidades recibirán en forma aleatoria los niveles del factor, mientras que las covariables son mediciones de una o más variables continuas tomadas en cada unidad experimental antes o durante la aplicación de los tratamientos (Quinn y Keough, 2002).

2.3. Estimación de parámetros.

A continuación se da una descripción de los métodos más utilizados para la estimación de parámetros en modelos lineales generales, los que corresponden a; mínimos cuadrados ordinarios

(OLSE, por sus siglas en inglés), mínimos cuadrados ponderados (WLSE o GLSE, por sus siglas en inglés) y máxima verosimilitud (MLE, por sus siglas en inglés) (Searle, 1987).

2.3.1. Estimaciones de μ y β . Los parámetros en un modelo de efectos fijos se pueden estimar usando el método de mínimos cuadrados ordinarios (MCO) (Sanni y Ukaegbu, 2012), el cual es un método no iterativo en que se asume que \mathbf{y} es un vector de variables aleatorias con una distribución normal que tiene media $\mathbf{X}\beta$ y varianza $\mathbf{I}\sigma^2$. Para calcular las soluciones del vector β se utilizan ecuaciones con la siguiente nomenclatura matricial (Kaps y Lamberson, 2009).

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{y}$$

Formula 4. Ecuaciones de mínimos cuadrados ordinarios.

Si la matriz \mathbf{X} tiene rango completo, entonces $\mathbf{X}'\mathbf{X}$ tiene inversa, y la solución a la ecuación anterior se obtiene mediante $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Sin embargo, con muchas formas de modelos lineales, como en el caso del modelo de los efectos, la matriz \mathbf{X} no tiene rango completo, lo que implica que $\mathbf{X}'\mathbf{X}$ no tiene inversa. En tales casos, se usa una inversa generalizada \mathbf{G} para obtener las soluciones de β mediante la expresión $\beta = \mathbf{G}\mathbf{X}'\mathbf{y}$ (Searle, 1987).

No obstante, debe tenerse presente que cuando $\mathbf{X}'\mathbf{X}$ tiene inversa se obtiene una única solución para el vector β , pero cuando $\mathbf{X}'\mathbf{X}$ no tiene inversa, la solución para el vector de parámetros β es sólo una de un infinito número de soluciones posibles para las ecuaciones normales (Shaw y Mitchell – Olds, 1993).

Este método funciona de manera óptima en modelos lineales con un único término de error, y bajo el cumplimiento del supuesto de homogeneidad de varianza. No obstante, en modelos lineales mixtos, el MCO ya no presenta un funcionamiento óptimo (Piepho *et al.*, 2003).

2.3.2. Mínimos cuadrados ponderados. La inferencia basada en mínimos cuadrados ordinarios (MCO) tiene como uno de los supuestos básicos a la homogeneidad de varianzas (Abdi, 2003), también conocida como homocedasticidad (Miranda, 2011b), que en caso de no cumplirse hace que el estimador de MCO sea insesgado y consistente pero ineficiente, y bajo estas condiciones, los MCO ya no son el mejor estimador lineal e insesgado (Ramírez y Ramírez, 2006). Además, la heterocedasticidad invalida las inferencias (pruebas t, pruebas F, etc.) (Sosa, 2009).

Uno de los métodos alternativos de análisis, que se utiliza en caso de heterocedasticidad, es el método de mínimos cuadrados ponderados (MCP) (Sosa, 2009), de los cuáles los MCO son un caso especial (Abdi, 2003; Knaub, 2010), y que es capaz de proporcionar el mejor estimador linealmente insesgado, pero que depende de conocer la estructura de varianza, lo que en la práctica raramente está disponible (Ramírez y Ramírez, 2006).

2.3.3. Máxima verosimilitud. Máxima verosimilitud, o ML por sus siglas en inglés, es un método iterativo destinado a la estimación de parámetros (μ y σ^2), tanto en modelos de efectos fijos como en modelos de efectos aleatorios y modelos mixtos (Kaps y Lamberson, 2009).

En este método de estimación se asume que los términos de error y los efectos aleatorios están normalmente distribuidos (Cadena y Castillo, 2000).

2.4. Estimación de componentes de varianza.

En esta sección se describen métodos para la estimación de componentes de varianza, que se utilizan en el ámbito agropecuario para el mejoramiento genético animal o vegetal, y mediante los cuáles es posible determinar el coeficiente de correlación intraclase e índices de heredabilidad y repetibilidad, respuesta a la selección, coeficientes de diferenciación genética y de variabilidad genética aditiva, entre otros (Rasch y Masata, 2006; Di Rienzo *et al.*, 2011).

2.4.1. Máxima verosimilitud restringida. El método iterativo de máxima verosimilitud restringida (Littell *et al.*, 1996), también conocido como máxima verosimilitud residual (Bates y Pinheiro, 2001; Bates, 2008), está destinado a la estimación de componentes de varianza (Solomon, 2005). Se puede utilizar tanto en modelos de efectos aleatorios como en modelos mixtos (Kaps y Lamberson, 2009), y cuenta con un amplio uso y aceptación actualmente, reemplazando a otros métodos más antiguos como el método de momentos (SAS Institute Inc., 2010a), especialmente porque evita la determinación de componentes de varianza negativos (Cadena y Castillo, 2000).

REML es una variante de ML (Rasch y Masata, 2006), que también asume una distribución normal, pero que a diferencia de ML no provee estimaciones de $\mathbf{X}\beta$, aunque si toma en cuenta los grados de libertad asociados con la estimación de los efectos fijos (Searle, 1987; Ramroop, 2002).

2.4.2. MINQUE y MIVQUEO. Estos métodos utilizados para la estimación de componentes de varianza son recomendados principalmente para muestras grandes (Goodnight, 1978; Solomon, 2005; Junhuan *et al.*, 2011), donde MINQUE, a diferencia de ML y REML, no requiere que se cumpla el supuesto de normalidad. Cabe señalar que tanto MINQUE como MIVQUEO corresponden a métodos no iterativos (Searle, 1987; Littell *et al.*, 1996).

2.5. Predicción de factores aleatorios.

Para obtener una predicción del vector de efectos aleatorios \mathbf{u} en un modelo tipo II se pueden utilizar ecuaciones de mínimos cuadrados generalizados (Kaps y Lamberson, 2009).

En el caso de un modelo tipo III, se puede recurrir al uso de ecuaciones de modelos mixtos para obtener una predicción del vector de efectos aleatorios \mathbf{u} , las cuáles son ecuaciones normales extendidas. Con las ecuaciones de modelos mixtos también se puede obtener una estimación del vector de efectos fijos β (Littell *et al.*, 1996; Kaps y Lamberson, 2009).

2.6. Estimación de sumas de cuadrados.

En una tabla de análisis de varianza de Fisher, sus columnas corresponden a; fuentes de variación, grados de libertad, suma total de cuadrados, cuadrados medios, valores F y significancia o error alfa (Miranda, 2011b).

En relación con las sumas de cuadrados (SS), éstas descomponen la variabilidad total, atribuible a la variable de respuesta, en el o los factores controlados y el error experimental, que constituye la fuente de variación que aglutina a todos los componentes no controlados (Restrepo, 2007).

En el caso más sencillo, correspondiente a un ANOVA de una vía, tres fuentes de variabilidad son definidas y medidas, con sus sumas de cuadrados respectivas, las cuáles corresponden a las sumas de cuadrados total, sumas de cuadrados entre grupos o tratamientos (variabilidad explicada) y sumas de cuadrados dentro de grupos o tratamientos (variabilidad no

explicada), ésta última también conocida como suma de cuadrados residual o suma de cuadrados del error (Kaps y Lamberson, 2009; Spiegel *et al*, 2010).

Se debe tener en cuenta que el objetivo tradicional de los diseños experimentales corresponde a minimizar la suma de cuadrados del error (residual), para así maximizar el rechazo de la hipótesis nula, y de este modo poder detectar diferencias entre tratamientos (Restrepo, 2007).

Dentro de los diferentes tipos de sumas de cuadrados, se encuentran las sumas de cuadrados tipo I, II, III y IV.

2.6.1. Sumas de cuadrados tipo I. Las sumas de cuadrados tipo I, también conocidas como método de descomposición jerárquica de las sumas de cuadrados (IBM, 2002), o sumas de cuadrados secuenciales, dependen del orden de los términos en el modelo, y en ellas, al sumarse las sumas de cuadrados de los efectos individuales se obtiene como resultado la suma de cuadrados total o suma de cuadrados del modelo (Hector *et al.*, 2010a), característica que no posee ninguna otra suma de cuadrado, excepto en casos especiales (SAS Institute Inc., 2009).

En este tipo de sumas de cuadrados, si se tienen diferentes ordenes en los términos del modelo se pueden obtener resultados muy distintos (Langsrud, 2003).

En las sumas de cuadrados secuenciales, cada término se corrige sólo respecto al término que le precede en el modelo (IBM, 2002).

El método SS Tipo I para la obtención de sumas de cuadrados se utiliza normalmente para un modelo de ANOVA equilibrado especificado en la secuencia apropiada, es decir, un modelo en el que se especifica cualquier efecto principal antes de cualquier efecto de interacción de primer orden, cualquier efecto de interacción de primer orden se especifica antes de cualquier efecto de interacción de segundo orden, y así sucesivamente. También se puede utilizar en un modelo puramente anidado, en donde el primer efecto especificado está anidado dentro del segundo efecto especificado, el segundo efecto especificado está anidado dentro del tercero, y así sucesivamente (Littell *et al.*, 1991; IBM, 2002).

Los resultados de las sumas de cuadrados secuenciales coinciden con la suma de cuadrados tipo II, III y IV para diseños balanceados (Restrepo, 2007).

2.6.2. Sumas de cuadrados tipo II. Este tipo de sumas de cuadrados, a diferencia de las sumas de cuadrados tipo I, no dependen del orden de los términos en el modelo (SAS Institute Inc., 2009). Sin embargo, los resultados para las sumas de cuadrados tipo II pueden ser obtenidos realizando varios análisis tipo I en distinto orden (Langsrud, 2003).

Este método calcula cada suma de cuadrados del modelo considerando sólo los efectos pertinentes, siendo estos los correspondientes a todos los efectos que no contienen el que se está examinando (IBM, 2002). En otras palabras, la suma de cuadrados tipo II para un efecto dado, pudiendo ser éste un efecto principal o una interacción, es ajustada para otro efecto siempre y cuando este segundo efecto no contenga al primero. Por ejemplo, en el caso de un experimento de dos factores con interacción, los efectos principales A y B no son ajustados para la interacción AB, debido a que contiene tanto al efecto A como al efecto B, mientras que el factor A es ajustado para B, ya que B no contiene a A. Similarmente, B es ajustado para A, y la interacción AB es ajustada para los dos efectos principales (Littell *et al.*, 1991).

El método SS Tipo II para la obtención de sumas de cuadrados se utiliza normalmente para; un modelo de ANOVA equilibrado (debido a que depende de las frecuencias de celdas), cualquier modelo que sólo tenga efectos de factor principal (sin interacción) o un diseño puramente anidado (IBM, 2002).

Las sumas de cuadrados tipo II coinciden con las sumas de cuadrados parciales, tipo III y IV, si no existe interacción en el modelo (Searle, 1987; Restrepo, 2007).

2.6.3. Sumas de cuadrados tipo III. Al igual que las sumas de cuadrados tipo II, y a diferencia de las sumas de cuadrados tipo I, las sumas de cuadrados tipo III no dependen del orden de los términos en el modelo (SAS Institute Inc., 2009). Sin embargo, los resultados para las sumas de cuadrados tipo II y III pueden ser obtenidos realizando varios análisis tipo I en distinto orden (Langsrud, 2003).

Son sumas de cuadrados parciales, y en este sentido, cada efecto es ajustado para todos los otros efectos (Littell *et al.*, 1991).

Las sumas de cuadrados Tipo III tienen una gran ventaja por ser invariables respecto a las frecuencias de casilla, a diferencia de las sumas de cuadrados tipo I y II (Littell *et al.*, 1991). Así, este tipo de sumas de cuadrados se suele considerar de gran utilidad para un modelo no equilibrado sin casillas perdidas (IBM, 2002).

El método SS Tipo III para la obtención de sumas de cuadrados se utiliza normalmente para cualquiera de los modelos que aparecen en los tipos I y II o cualquier modelo equilibrado o

no, pero sin casillas vacías (Hector *et al.*, 2010b). Su uso principal es en situaciones que requieren una comparación de efectos principales incluso en presencia de interacción (Littell *et al.*, 1991; Langsrud, 2003).

2.6.4. Sumas de cuadrados tipo IV. Las sumas de cuadrados tipo IV, al igual que las sumas de cuadrados tipo III, a menudo se denominan sumas de cuadrados parciales (SAS Institute Inc., 2009).

El método SS Tipo IV para la obtención de sumas de cuadrados se utiliza normalmente para cualquiera de los modelos que aparecen en los tipos I y II o cualquier modelo equilibrado o no que tenga casillas vacías (IBM, 2002; Littell *et al.*, 1991; Hector *et al.*, 2010b).

Si todas las celdas están ocupadas, las sumas de cuadrados tipo III y IV coinciden (Searle, 1987). Si al menos una celda está vacía, las sumas de cuadrados tipo III y IV son distintas, y se asocian a diferentes hipótesis (Restrepo, 2007). Ante la presencia de celdas vacías la suma de cuadrados tipo IV puede no ser única (Littell *et al.*, 1991), ya que depende de la posición y del número de celdas vacías (Restrepo, 2007).

2.7. Evaluación del ajuste de los modelos.

A continuación se describirán dos métodos habitualmente empleados para la evaluación del ajuste de modelos estadísticos a un conjunto de datos. Tales métodos son los criterios de información de Akaike y Bayesiano, que son los más comúnmente usados para el propósito de evaluar ajustes de modelos (de – Graft, 2010).

2.7.1. Criterio de información de Akaike (AIC). La base teórica del método de Akaike combina; teoría de máxima verosimilitud, teoría de la información (información Kullback – Leibler) y el concepto de entropía de la información (entropía de Boltzman) (Motulsky y Christopoulos, 2003; Posada y Rosero, 2007).

El AIC ofrece una medida relativa de la información perdida cuando un modelo es usado para describir la realidad. Respecto a lo anterior, si se tienen datos generados por un proceso desconocido f , y se tienen dos modelos candidatos para representar a f : g_1 y g_2 , en el caso de conocerse f se podría encontrar la información perdida por el uso de g_1 y g_2 para representar a f , mediante el cálculo de la divergencia Kullback – Leibler, eligiéndose el modelo candidato que minimice la información perdida, el cuál tendrá un menor valor K – L. Sin embargo, no se puede elegir con certeza, debido a que se desconoce f , pero se puede minimizar la estimación de la información perdida (Burnham y Anderson, 2004; Posada y Buckley, 2004; Liddle, 2008; de – Graft, 2010; Caballero, 2011; Hu, 2012).

El método de Akaike puede ser usado para comparar tanto modelos anidados como no anidados (Liddle, 2008; Hu, 2012), ajustados a un mismo conjunto de datos (Bates y Pinheiro, 2001), y permite determinar cuál modelo es más probable que sea correcto, y cuantificar cuanto más probable es (Motulsky y Christopoulos, 2003).

La ecuación para el cálculo del AIC se muestra a continuación. Donde $l_{m\acute{a}x}$ es el logaritmo de verosimilitud maximizado y k es el número de parámetros, siendo el mejor modelo aquel que tenga el menor valor de AIC (Posada y Buckley, 2004; Posada y Rosero, 2007; Liddle, 2008; de – Graft, 2010; Hu, 2012).

$$AIC = -2 l_{m\acute{a}x} + 2 k$$

Fórmula 5. Criterio de Información de Akaike.

En la ecuación anterior, el primer término es una medida de la bondad de ajuste, y el segundo es un término de penalización del número de parámetros (Caballero, 2011). No obstante, cabe señalar que no es necesario usar aquella ecuación (Motulsky y Christopoulos, 2003), sólo se debe elegir el modelo con menor valor AIC, ya que la utilidad de este criterio está en la comparación de modelos, por lo tanto, la diferencia entre valores AIC es lo que importa (Posada y Buckley, 2004).

Antiguamente se empleaba este criterio en el sentido inverso, y por lo tanto un mayor valor numérico correspondía a un mejor ajuste (Bates y Pinheiro, 2001; Bates, 2008).

Criterio de información de Akaike corregido (AIC_C). Cuando N no es suficientemente grande comparado con K el AIC es demasiado pequeño, por lo que se ha desarrollado el AIC corregido que es más exacto, y que se calcula con la siguiente ecuación (Caballero, 2011; Hu, 2012).

$$AIC_C = AIC + (2 K (K + 1)) / (N - K - 1)$$

Fórmula 6. Criterio de Información de Akaike Corregido.

Si las muestras son grandes, con al menos unas pocas docenas de veces más puntos de datos que parámetros, esta corrección es innecesaria. Con muestras pequeñas, la corrección será de importancia y ayudará a elegir el mejor modelo. Al igual que con el AIC, el modelo con un menor AIC_C es el que tiene más probabilidad de ser correcto (Motulsky y Christopoulos, 2003).

Posada y Buckley (2004), recomiendan usar el AIC_C si $n / k < 40$, sin embargo, Motulsky y Christopoulos (2003) y Burnham y Anderson (2004) señalan que alternativamente se podría usar siempre el AIC_C en lugar del AIC.

2.7.2. Criterio de Información Bayesiano (BIC). AIC y BIC son probablemente los criterios más comúnmente usados para selección de modelos (Caballero, 2011), y en ambos, los valores más bajos representan mejores ajustes (Quinn y Keough, 2002; Burnham y Anderson, 2004; Posada y Rosero, 2007; de – Graft, 2010).

El BIC está definido por la siguiente ecuación, donde K es el número de parámetros estimables y n es el tamaño de la muestra (Posada y Buckley, 2004).

$$BIC = -2l + K \log n$$

Fórmula 7. Criterio de Información de Bayesiano.

En comparación con AIC, BIC tiende a penalizar los modelos complejos más fuertemente, dando preferencia a los modelos más simples (Quinn y Keough, 2002). Además, el enfoque del BIC, a diferencia del AIC, supone que el modelo real se encuentra incluido entre los modelos candidatos que se están evaluando (Caballero, 2011).

3. MATERIALES Y MÉTODOS.

3.1 Materiales.

En esta sección se describen los ejemplos seleccionados para realizar los análisis estadísticos, los que se llevaron a cabo con; la función lmer, contenida en el paquete lme4 del programa estadístico R, el procedimiento PROC MIXED del programa estadístico SAS y el procedimiento Fit Model del programa estadístico JMP, siendo estos dos últimos programas productos comerciales de la compañía SAS Institute Inc.

Se seleccionaron ejemplos que estuvieran disponibles en Internet, basados en resultados obtenidos mediante su análisis con SAS.

Cada uno de los análisis fue desarrollado utilizando los tres programas computacionales de análisis estadístico, con el objeto de realizar la evaluación de la similitud entre los resultados obtenidos con cada uno de ellos.

3.1.1. Diseño de un factor con bloque y covarianza. El ejemplo de ANOVA de un factor con bloque y covarianza que se presenta a continuación se obtuvo de Littell *et al* (1996), y los datos se encuentran disponibles en <http://ftp.sas.com/samples/A55235>.

La tabla de ANOVA de Fisher corresponde a la siguiente.

Cuadro 1. Tabla de ANOVA teórica bloques al azar.

Fuente de Variación	Grados de Libertad
Bloque	r-1
Cova	a-1
A	b-1
Error residual	(a-1)*(r-1)

En el ejemplo seleccionado se midió la ganancia promedio de peso diaria de 32 novillos (GPD), los cuáles fueron alimentados con cuatro dietas (correspondientes a los niveles del factor), que se componen de una ración base más tres niveles de alimento suplementario adicionado a la ración base, y suministrado por un período de 160 días. El objetivo del experimento es determinar el nivel óptimo de alimento suplementario para maximizar la ganancia de peso. Los novillos fueron ubicados en ocho graneros, que actuaban como bloques, asignándose las raciones a los novillos en forma aleatoria. Debido a que los novillos tenían distintos pesos iniciales este fue tomado como una posible covariable (PI).

Los datos son descritos por el modelo $y_{ij} = \alpha_i + \beta_i x_{ij} + b_j + \varepsilon_{ij}$. Donde $i = 1, 2, 3, 4$ y $j = 1, 2, \dots, 8$; α_i denota el intercepto de la i -ésima dieta; β_i denota la pendiente de la i -ésima dieta; b_j denota el efecto del i -ésimo bloque, con distribución $N(0, \sigma_b^2)$ y ε_{ij} denota el error experimental, con distribución $N(0, \sigma_\varepsilon^2)$, siendo b_j y ε_{ij} variables aleatorias independientes.

La tabla de datos para el ejemplo, que se adjunta en la sección de anexos, corresponde a un conjunto de datos balanceados, y tiene la siguiente estructura.

Cuadro 2. Estructura de la tabla de datos para el ejemplo de un factor con bloque y covarianza.

Bloque	Tratamiento	GPD	PI
B_1	T_0	1,03	338
B_1	T_10	1,54	477
B_1	T_20	1,82	444
...			
B_8	T_10	0,64	376
B_8	T_20	0,76	308
B_8	T_30	0,7	439

No obstante, se realizó un procedimiento de detección de outliers eliminándose uno de los datos más extremos (dato n° 30 de la tabla), y se procedió a realizar el análisis sin aquel dato.

3.1.2. Diseño de un factor fijo y un factor aleatorio. El ejemplo de ANOVA de dos factores que se presenta a continuación se obtuvo de la página web de SAS Institute Inc., y los datos se encuentran disponibles en http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_glm_sect054.htm.

La tabla de ANOVA de Fisher corresponde a la siguiente.

Cuadro 3. Tabla de ANOVA teórica dos factores Completamente Aleatorizado.

Fuente de Variación	Grados de Libertad
A	a-1
B	b-1
A*B	(a-1)*(b-1)
Error residual	a*b*(r-1)

En este ejemplo se tienen tres máquinas y seis operarios. Cada uno de los operarios utiliza cada máquina en una, dos o tres ocasiones. La variable dependiente es una calificación, que toma en cuenta el número y cantidad de componentes producidos.

Los datos son descritos por el modelo $y_{ijk} = \mu + \tau_i + \eta_j + (\tau\eta)_{ij} + \varepsilon_{ij}$. Donde y_{ijk} es la k -ésima observación, en la i -ésima máquina utilizada por el j -ésimo operario; μ es el intercepto; τ_i es el efecto fijo de la i -ésima máquina; η_j es el efecto aleatorio del j -ésimo operario; $(\tau\eta)_{ij}$ es la ij -ésima interacción entre máquina y operario; mientras que ε_{ij} es el error aleatorio.

El operario, la interacción entre máquina y operario, y el error residual, se asume que están normalmente distribuidos, con media 0 y componentes de varianza σ_O^2 , σ_{MO}^2 y σ^2 , respectivamente.

La tabla de datos para el ejemplo, que se adjunta en la sección de anexos, corresponde a un conjunto de datos desbalanceado, y tiene la siguiente estructura.

Cuadro 4. Estructura de la tabla de datos para el ejemplo de un factor fijo y un factor aleatorio.

Maquina	Operario	Calificación
M_1	E_1	52
M_1	E_2	51,8
M_1	E_2	52,8
...		
M_3	E_6	62
M_3	E_6	61,4
M_3	E_6	60,5

3.1.3. Diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.

El ejemplo de ANOVA de dos factores que se presenta a continuación se obtuvo de Littell *et al* (1996), y los datos se encuentran disponibles en <http://ftp.sas.com/samples/A55235>.

La tabla de ANOVA de Fisher corresponde a la siguiente.

Cuadro 5. Tabla de ANOVA teórica Parcelas Dividida en Bloque completos al azar.

Fuente de Variación	Grados de Libertad
Bloque	$r-1$
A	$a-1$
Error	$(r-1) * (a-1)$
B	$b-1$
A*B	$(a-1) * (b-1)$
Error residual	$a * (r-1) * (b-1)$

En este ejemplo, se comparan dos tratamientos en cada una de siete granjas.

Los datos son descritos por el modelo $y_{ijk} = \mu + F_i + R(F)_{ij} + \tau_k + FT_{ik} + \varepsilon_{ijk}$. Donde y_{ijk} es la observación en el j –ésimo bloque, i –ésima granja y k –ésimo tratamiento; μ es el intercepto, F_i es el efecto de la i –ésima granja; $R(F)_{ij}$ es el efecto de bloque dentro de granja; τ_k es el efecto fijo del k –ésimo tratamiento; FT_{ik} es el efecto de interacción de granja por tratamiento y ε_{ijk} es el error aleatorio. Debido a que las granjas son una muestra aleatoria, los efectos F_i , $R(F)_{ij}$ y FT_{ik} se asume que son aleatorios, con distribuciones; $F_i \sim iid N(0, \sigma_F^2)$, $R(F)_{ij} \sim iid N(0, \sigma_R^2)$ y $FT_{ik} \sim iid N(0, \sigma_{FT}^2)$.

La tabla de datos para el ejemplo, que se adjunta en la sección de anexos, corresponde a un conjunto de datos balanceados, y tiene la siguiente estructura.

Cuadro 6. Estructura de la tabla de datos para el ejemplo de un factor fijo y un factor aleatorio.

Granja	Repetición	Tratamiento	Y
G_1	R_1	T_1	49,7
G_1	R_1	T_2	56,2
G_1	R_2	T_1	41
...			
G_7	R_1	T_2	40,2
G_7	R_2	T_1	37,8
G_7	R_2	T_2	31,9

Para la obtención de un diseño desbalanceado se realizó un procedimiento de detección de outliers, decidiendo eliminarse el dato más extremo (dato n° 26).

3.1.4 Diseño de parcelas divididas. El ejemplo que se presenta a continuación corresponde a un diseño de parcelas divididas, que fue extraído del sitio web de SAS Intitute Inc. y se encuentra disponible en http://support.sas.com/documentation/cdl/en/statug/65328/HTML/default/viewer.htm#statug_mixed_examples01.htm.

La tabla de ANOVA de Fisher corresponde a la siguiente.

Cuadro 7. Tabla de ANOVA teórica Parcelas Dividida en Bloque completos al azar.

Fuente de Variación	Grados de Libertad
Bloque	r-1
A	a-1
Error	(r-1)*(a-1)
B	b-1
A*B	(a-1)*(b-1)
Error residual	A*(r-1)*(b-1)

El diseño es balanceado y consta de cuatro bloques, donde los tres niveles del factor principal (factor A) se encuentran asignados de manera aleatoria dentro de cada bloque, mientras que los dos niveles del factor secundario (factor B) se encuentran aleatorizados dentro de cada parcela principal.

Los datos son descritos por el modelo $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + B_k + E_{ik} + \varepsilon_{ijk}$. Donde y_{ijk} es la observación en el k – *ésimo* bloque, i – *ésimo* nivel de riego y j – *ésima* variedad; μ es la media total; α_i es el efecto fijo del i – *ésimo* nivel de riego; β_j es el efecto fijo de la j – *ésima* variedad; $(\alpha\beta)_{ij}$ es la interacción del i – *ésimo* nivel de riego con la j – *ésima* variedad; B_k es el efecto del k – *ésimo* bloque; E_{ik} es el término de error para la parcela principal, dado por la interacción $B_k \times \alpha_i$; mientras que ε_{ijk} es el término de error para la subparcela.

Los términos de error y el efecto de bloque son efectos aleatorios, con media cero y varianzas σ^2_E , σ^2_v y σ^2_B , respectivamente.

La tabla de datos para el ejemplo, que se adjunta en la sección de anexos, corresponde a un conjunto de datos balanceados, y tiene la siguiente estructura.

Cuadro 8. Estructura de la tabla de datos para el ejemplo de parcelas divididas.

Bloque	Riego	Variedad	Y
B_1	R_1	V_1	56
B_1	R_1	V_2	41
B_1	R_2	V_1	50
...			
B_4	R_2	V_2	30
B_4	R_3	V_1	17
B_4	R_3	V_2	18

No obstante, se realizó un procedimiento de detección de outliers eliminándose el dato más extremo (dato n° 1 de la tabla), y se procedió a realizar el análisis sin aquel dato.

3.1.5. Diseño cross – over. El ejemplo de diseño crossover que se presenta a continuación se obtuvo de la publicación “Analysis of 2x2 cross – over trials using an integrated macro package”. Los datos, contenidos en la publicación antes mencionada, se encuentran disponibles en <http://www.sascommunity.org/sugi/SUGI91/Sugi-91-257%20Chang%20Strang.pdf>.

La tabla de ANOVA de Fisher corresponde a la siguiente.

Cuadro 9. Tabla de ANOVA teórica para parcelas divididas completamente al azar.

Fuente de Variación	Grados de Libertad
A	a-1
Error	a*(r-1)
B	b-1
Período	(a-1)*(b-1)
Error Residual	A*(r-1)*(b-1)

El diseño cross – over 2 x 2 se puede resolver como un diseño de parcelas divididas, aunque con una leve modificación, dado que el efecto del período sustituye al efecto de la interacción.

Los datos corresponden a un estudio de asma bronquial, donde se compara la eficacia de dos tratamientos (A y B) con respecto al volumen (en litros) de aire expirado en un segundo. Se

tiene dos secuencias (AB y BA), y cada uno de los 17 sujetos es sometido a los tratamientos en dos períodos consecutivos.

Los datos son descritos por el modelo $y_{ijk} = \mu + E_{ik} + \pi_j + \tau_{d(i,j)} + \lambda_{d(i,j-1)} + \varepsilon_{ijk}$. Donde y_{ijk} es la respuesta del k –ésimo sujeto en el período j –ésimo, bajo la secuencia i , ($i = 1, 2; j = 1, 2; k = 1, \dots, n_j$); μ es la media total; E_{ik} es el error aleatorio entre sujeto, con distribución $iid N(0, \sigma^2_E)$; π_j es el efecto del período j –ésimo, $\pi_1 + \pi_2 = 0$; $d(i, j)$ es el tratamiento asignado en el período j , bajo la secuencia i ; τ_i es el efecto de tratamiento directo ($i = 1, 2; \tau_1 + \tau_2 = 0$); λ_i es el efecto residual de los tratamientos ($\lambda_{d(1,0)} = 0; \lambda_1 + \lambda_2 = 0$); mientras que ε_{ijk} es el error aleatorio dentro de sujetos, que se asume $iid N(0, \sigma^2)$.

La tabla de datos para el ejemplo, que se adjunta en la sección de anexos, corresponde a un conjunto de datos desbalanceado y tiene la siguiente estructura.

Cuadro 10. Estructura de la tabla de datos para el ejemplo de diseño cross – over.

Secuencia	Sujeto	Tratamiento	Periodo	VAE
S_1	S_1	T_1	P_1	1,28
S_1	S_1	T_2	P_2	1,33
S_1	S_2	T_1	P_1	1,6
...				
S_2	S_17	T_1	P_2	3,23
S_2	S_18	T_2	P_1	1,16
S_2	S_18	T_1	P_2	1,25

3.2 Métodos.

Respecto a los métodos, en SAS se utilizó el procedimiento PROC MIXED, el cual ajusta modelos lineales mixtos del tipo $\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$ (SAS Institute Inc., 2010c), y que fue la primera herramienta desarrollada para el análisis de modelos mixtos en SAS, a la cual posteriormente se sumaron NLMIXED, HPMIXED y GLIMMIX (Kiernan *et al.*, 2012). No obstante, cabe señalar que antes de la aparición de PROC MIXED, PROC GLM fue el principal procedimiento de SAS para el análisis de modelos mixtos (Gil, 2001).

En lo que se refiere al programa estadístico gratuito R, se utilizó la función lmer, contenida dentro del paquete lme4, que además provee funciones para ajustar y analizar modelos lineales mixtos (LMMs), modelos lineales mixtos generalizados (GLMMs) y modelos mixtos no lineales (NLMs) (Bates *et al.*, 2012). La función lmer fue diseñada para ajustar modelos lineales mixtos, y tiene un alcance similar al procedimiento PROC MIXED de SAS (Bates, 2005; Bates, 2008).

Anteriormente, para el análisis de modelos mixtos en R se utilizaba la función lme, contenida dentro del paquete nlme (Bates y Pinheiro, 2001). El código del paquete nlme se escribió de tal manera que fuera compatible con S y S – Plus, las versiones comerciales de R, a diferencia de lme4. Además hay ligeras diferencias en la sintaxis de lme4 respecto de nlme, lo que implica diferencias, también ligeras, entre las sintaxis de las funciones lme y lmer (Cayuela, 2010).

Con respecto a JMP, se trata de un programa estadístico desarrollado por SAS Institute Inc., de funcionamiento similar al procedimiento GLM de SAS, y con capacidad para analizar modelos lineales con factores fijos, aleatorios o modelos mixtos (SAS Institute Inc., 2010a).

4. PRESENTACIÓN Y DISCUSION DE RESULTADOS.

En esta sección se muestran los resultados resumidos de los análisis realizados, junto con una breve descripción de cada uno de los ejemplos seleccionados.

Las estimaciones de componentes de varianza se obtuvieron mediante el método REML (máxima verosimilitud restringida), el cual es el método predeterminado para estimación de componentes de varianza en los programas SAS, R y JMP, así como en la mayoría de los programas estadísticos actuales (Piepho *et al.*, 2003).

En cuanto a los valores del estadístico F y las significancias para cada uno de los factores fijos, estos se determinaron (en los tres programas) mediante sumas de cuadrados tipo III, que presentan la ventaja de ser invariables respecto a las frecuencias de casilla, a diferencia de las sumas de cuadrados tipo I y II (Littell *et al.*, 1991).

En la sección de anexos se encuentran las sintaxis y salidas originales para cada uno de los análisis realizados en SAS, R y JMP, donde es posible encontrar más detalles de los resultados.

Cabe recordar que este trabajo de investigación tiene como propósito realizar las estimaciones de diferentes modelos de análisis de varianza habitualmente empleados en investigación agropecuaria, con la finalidad de evaluar el grado de similitud obtenido por medio de diferentes programas estadísticos empleados en investigación.

4.1. Diseño de un factor con bloque y covarianza.

La variable de respuesta corresponde a la ganancia diaria de peso. El factor comparativo es la dieta, con cuatro niveles, y se tiene un factor de bloqueo dado por los distintos graneros. Adicionalmente se tiene una covariable correspondiente al peso inicial de los novillos.

4.1.1. Resultados de los análisis.

Cuadro 11. Estimaciones de componentes de varianza.

Componentes de varianza	SAS PROC MIXED	R	JMP
Bloque	0,2251	0,226864	0,2250501
Residual	0,05234	0,068962	0,0523441

Se observa que las estimaciones de componentes de varianza obtenidas mediante SAS, R y JMP entregan resultados muy similares, tanto para los graneros, que actúan como factor de bloqueo, como para el error residual, que representa la variabilidad no explicada.

Cuadro 12. Valores F y probabilidades.

Valor F	SAS PROC MIXED	R	JMP
Tratamiento	9,91	10,632	9,9103
Peso inicial	10,75	13,348	10,4932

Probabilidad	SAS PROC MIXED	R	JMP
Tratamiento	0,0004	0,0002528888	0,0004
Peso inicial	0,0039	0,001689644	0,0041

Se observa una pequeña variación en los valores F y probabilidades para el efecto del tratamiento, que representa cuatro dietas distintas, así como para el peso inicial, utilizado como covariable, cuando se comparan las estimaciones de SAS o JMP con R. Sin embargo, existe una marcada similitud entre los valores numéricos para el estadístico F y significancias entre SAS y JMP.

Cuadro 13. Evaluación del ajuste de los modelos.

Criterios de información	SAS PROC MIXED	R	JMP
AIC	69,1	85,12	
AICC	69,7		101,6687
BIC	69,3	99,46	105,0086

Se observa que los valores de BIC son mayores que los de AIC y AICC tanto en R como en JMP, mientras que en SAS, los valores de AIC, AICC y BIC son muy similares.

4.2. Diseño de un factor fijo y un factor aleatorio.

La variable de respuesta es una calificación. El efecto de las máquinas es fijo y el de los operarios aleatorio, siendo la interacción entre ambos aleatoria.

4.2.1. Resultados de los análisis.

Cuadro 14. Estimaciones de componentes de varianza.

Componentes de varianza	SAS PROC MIXED	R	JMP
Operario	24,2571	22,45579	22,455765
Máquina*Operario	17,0791	14,23401	14,233991
Residual	0,8726	0,87087	0,8708687

Se observan estimaciones de componentes de varianza muy similares entre R y JMP, y una leve, aunque notoria, diferencia en las estimaciones obtenidas mediante SAS en comparación a R y JMP.

Cuadro 15. Valores F y probabilidades.

Valor F	SAS PROC MIXED	R	JMP
Máquina	16,72	19,966	19,9639
Probabilidad	SAS PROC MIXED	R	JMP
Máquina	0,0006	0,0003221849	0,0003

Se observan valores F y significancias prácticamente idénticas entre R y JMP, no obstante, se observa una discrepancia con los resultados obtenidos mediante SAS.

Cuadro 16. Evaluación del ajuste de los modelos.

Criterios de información	SAS PROC MIXED	R	JMP
AIC	188,0	193,9	
AICC	188,7		198,339
BIC	187,4	204,6	206,7739

Se observa que los valores de BIC son mayores que los de AIC y AICC tanto en R como en JMP, mientras que en SAS, los valores de AIC, AICC y BIC son muy similares.

4.3. Diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.

Se comparan dos tratamientos en cada una de siete granjas, donde los tratamientos son considerados fijos, mientras las granjas son consideradas efectos aleatorios.

4.3.1. Resultados de los análisis.

Cuadro 17. Estimaciones de componentes de varianza.

Componentes de varianza	SAS PROC MIXED	R	JMP
Repetición	1,5434	1,5434	1,5434202
Granja	7,3813	7,3818	7,3812991
Repetición(Granja)	1,4623	1,4622	1,4622946
Granja*Tratamiento	43,3507	43,3505	43,350676
Residual	15,2256	15,2256	15,225618

Se observa que las estimaciones de componentes de varianza obtenidas mediante SAS, R y JMP entregan resultados muy similares, siendo prácticamente idénticas.

Cuadro 18. Valores F y probabilidades.

Valor F	SAS PROC MIXED	R	JMP
Tratamiento	2,96	2,9637	2,9235
Probabilidad	SAS PROC MIXED	R	JMP
Tratamiento	0,1359	0,1359349	0,1420

Se observan valores muy similares entre los tres programas, aunque más similares entre SAS y R, donde se obtuvieron resultados prácticamente idénticos.

Cuadro 19. Evaluación del ajuste de los modelos.

Criterios de información	SAS PROC MIXED	R	JMP
AIC	167,2	171,2	
AICC	170,7		179,1874
BIC	162,7	179,7	181,1313

Se observa que los valores de BIC son mayores que los de AIC y AICC tanto en R como en JMP, mientras que en SAS, los valores de AIC, AICC y BIC presentan cierta heterogeneidad.

4.4. Diseño de parcelas divididas.

A es el factor asignado a la parcela principal y corresponde a tres niveles de riego, y el factor B, asignado a las subparcelas, representa dos variedades de trigo.

4.4.1. Resultados de los análisis.

Cuadro 20. Estimaciones de componentes de varianza.

Componentes de varianza	SAS PROC MIXED	R	JMP
Bloque	50,9788	50,9938	50,993788
Riego*Bloque	12,9034	12,9032	12,903241
Residual	6,6968	6,6963	6,6963352

Se observan diferencias muy leves en las estimaciones de componentes de varianza obtenidas mediante los tres programas, habiendo mayor similitud entre R y JMP, donde se obtienen resultados prácticamente idénticos.

Cuadro 21. Valores F y probabilidades.

Valor F	SAS PROC MIXED	R	JMP
Riego	4,53	4,3987	4,5326
Variedad	17,66	17,2661	17,4549
Riego*Variedad	4,13	4,1317	4,1272
Probabilidad	SAS PROC MIXED	R	JMP
Riego	0,0631	0,06666494	0,0617
Variedad	0,0030	0,003185921	0,0029
Riego*Variedad	0,0586	0,05854834	0,0572

Se observan resultados muy similares entre los tres programas, obteniéndose resultados prácticamente idénticos para el factor riego entre SAS y JMP, y para la interacción entre riego y variedad entre SAS y R.

Cuadro 22. Evaluación del ajuste de los modelos.

Criterios de información	SAS PROC MIXED	R	JMP
AIC	115,3	127,3	
AICC	117,1		149,6885
BIC	113,4	137,5	146,0618

Se observa que los valores de BIC son mayores que los de AIC en R, mientras que AICC es mayor que BIC en JMP. Por su parte, en SAS los valores de AIC, AICC y BIC presentan cierta heterogeneidad.

4.5. Diseño cross – over.

Se tienen dos tratamientos (A y B) y dos secuencias (AB y BA). Cada uno de los 17 sujetos es sometido a los tratamientos en dos períodos consecutivos.

4.5.1. Resultados de los análisis.

Cuadro 23. Estimaciones de componentes de varianza.

Componentes de varianza	SAS PROC MIXED	R	JMP
Sujeto(Secuencia)	0,3618	0,36178	0,3617763
Residual	0,1192	0,11921	0,1192124

Se observa que las estimaciones de componentes de varianza obtenidas mediante SAS, R y JMP entregan resultados muy similares, siendo prácticamente idénticas.

Cuadro 24. Valores F y probabilidades.

Valor F	SAS PROC MIXED	R	JMP
Secuencia	2,64	2,6347	2,6357
Tratamiento	4,68	4,9960	4,6759
Período	1,37	1,3734	1,3734
Probabilidad	SAS PROC MIXED	R	JMP
Secuencia	0,1253	0,1253759	0,1253
Tratamiento	0,0472	0,04103929	0,0472
Período	0,2595	0.2595068	0,2595

Se observan resultados similares entre los tres programas, no obstante, las diferencias son menores entre SAS y JMP, obteniéndose resultados prácticamente exactos.

Cuadro 25. Evaluación del ajuste de los modelos.

Criterios de información	SAS PROC MIXED	R	JMP
AIC	64,6	72,61	
AICC	65,1		79,8772
BIC	66,3	81,77	85,92425

Se observa que los valores de BIC son mayores que los de AIC y AICC tanto en R como en JMP y SAS. Por su parte, en SAS los valores de AIC, AICC y BIC presentan cierta heterogeneidad.

Cuadro 26. Cuadro comparativo de los criterios de información de Akaike y Bayesiano en SAS, R y JMP.

Diseño	SAS PROC MIXED			R		JMP	
	AIC	AICC	BIC	AIC	BIC	AICC	BIC
3.1.1.	40,5	41,0	40,7	55,9	63,07	58,14163	63,30997
3.1.2.	188,0	188,7	187,4	193,9	204,6	198,339	206,7739
3.1.3.	167,2	170,7	162,7	171,2	179,7	179,1874	181,1313
3.1.4.	115,3	117,1	113,4	127,3	137,5	149,6885	146,0618
3.1.5.	64,6	65,1	66,3	72,61	81,77	79,8772	85,92425

Se observa que en cada uno de los diseños experimentales los valores numéricos de los criterios de información de Akaike y Bayesiano fueron menores en SAS y más altos en JMP, con R presentando valores numéricos de AIC y BIC intermedios respecto a SAS y JMP. Estas diferencias se atribuyen a que cada programa estadístico calcula de manera distinta los parámetros del modelo.

5. CONCLUSIONES.

Luego de haber realizado los análisis de los diseños experimentales seleccionados, se comprueba que los programas utilizados, tanto gratuitos como comerciales, poseen características interesantes para su aplicación en ciencias agropecuarias, y a pesar de que tanto en R como en SAS existe la necesidad de codificar las sintaxis de cada procedimiento estadístico, esta es una labor abordable. No obstante, se destaca la facilidad de uso del programa JMP, que no requiere codificación de sintaxis en comparación con SAS y R.

Sobre la base del trabajo realizado, se confirma que en general hubo coincidencia en los resultados obtenidos con los tres programas estadísticos, llegando a obtenerse de manera frecuente coincidencias exactas entre sus resultados, tanto en las estimaciones de componentes de varianza para los efectos aleatorios, como en los valores de los estadísticos F y probabilidades para los efectos fijos, y en las estimaciones de las medias de cuadrados mínimos.

Una diferencia importante de destacar se refiere a que, para cada uno de los diseños experimentales analizados, los valores numéricos de los criterios de información fueron menores en el programa estadístico SAS, a su vez, los valores numéricos de los criterios de información en R fueron mayores a los de SAS, mientras que JMP presentó valores mayores que R para los criterios de información.

Finalmente, en relación con la facilidad de uso y presentación de la salida de resultados, los programas estadísticos se pueden ordenar, de mayor a menor facilidad de uso y organización en el despliegue del listado de sus resultados en; JMP, SAS y R, lo cual debe ser considerado en relación a su uso a nivel de pregrado.

6. RESUMEN.

Se llevó a cabo una comparación de la bondad de ajuste global, por medio de los criterios de información de Akaike y Bayesiano, entre el programa gratuito R de CRAN, utilizando en los análisis el paquete lme4, y los procedimientos MIXED y Fit Model de los programas comerciales SAS y JMP respectivamente.

Para dicha evaluación se recurrió a diseños experimentales de uso habitual en ciencias agropecuarias, los que permitieron evaluar los programas computacionales respecto a factores fijos y aleatorios, factores de bloqueo y covarianza, en un contexto de datos no balanceados. Lo anterior se realizó mediante la aplicación de dos variaciones de diseños de parcelas divididas, un modelo mixto full factorial, y un diseño de una vía con covarianza y factor de bloqueo. Los datos utilizados están disponibles en Internet, y su ubicación se encuentra debidamente citada, con el fin de facilitar la replicación de los análisis por parte de los lectores interesados en ello.

Las sintaxis, así como las salidas para efectos fijos y aleatorios se presentan en tablas, y se adicionan los resultados de las medias de cuadrados mínimos y Criterios de Información de Akaike y Bayesiano.

Sobre la base del trabajo realizado, una conclusión con implicancias prácticas consiste en la obtención de resultados similares mediante el software gratuito R, llegando a obtenerse de manera frecuente coincidencias exactas, tanto en las estimaciones de componentes de varianza, como en los valores de los estadísticos F y significancias, y en las medias de cuadrados mínimos.

Una diferencia importante de destacar se refiere a que, para cada uno de los diseños experimentales analizados, los valores numéricos de los criterios de información fueron menores en el programa estadístico SAS, a su vez, los valores numéricos de los criterios de información en R fueron mayores a los de SAS, mientras que JMP presentó valores mayores que R para los criterios de información.

7. SUMMARY.

A comparison was made of the goodness of fit, using the Akaike Information Criteria and Bayesian Information Criteria, in three statistical analysis computer programs, one free and two commercial. The free program selected for this work was R, assessed using the lme4 package, which is compared with the MIXED and Fit Model procedures from commercial programs SAS and JMP respectively.

For this assessment was used experimental designs frequently used in agricultural sciences. The selected designs allowed us to evaluate computational programs for fixed and random factors, blocking factors and covariance, in a context of unbalanced data sets. This was done by applying two variations of split plot design, a full factorial mixed model, and an design of one – way with block and covariance.

The data used are available on the Internet, and are properly cited, to facilitate analysis replication by the readers interested.

The syntax and outputs are presented in tables, and are added results of the least squares means and Akaike and Bayesian Information Criterion.

A conclusion with practical implications is to obtain similar results using free R, comparing with the other softwares. Very often, exact match for variance component estimates or for fixed effects estimates was reached between softwares, and for the least square means.

A noteworthy difference means that, for each of the experimental designs analyzed, the numerical values of the information criteria were lower in the SAS statistical program. The numerical values of the information criteria in R were higher than those of SAS, while JMP presented higher values than R to the information criterias.

8. LITERATURA CITADA.

- Abdi, H.** 2003. Least Squares. Disponible en <http://www.utdallas.edu/~herve/Abdi-LeastSquares-pretty.pdf>. Consultado el 22 de abril de 2013.
- Bates, D y Pinheiro, J.** 2001. lme for SAS PROC MIXED Users. Disponible en http://www.google.cl/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDAQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.130.1698%26rep%3Drep1%26type%3Dpdf&ei=EQeAUcayE6ndyAGIwIHQDA&usg=AFQjCNHGnfS8l9o9i_shfG3Ru7rk5OTFCQ&bvm=bv.45645796,d.aWc. Consultado el 8 de octubre de 2012.
- Bates, D.** 2005. Fitting linear mixed models in R. R News (Estados Unidos). 5 (1): 27 – 30. Disponible en http://mirrors.dotsrc.org/pub/cran/doc/Rnews/Rnews_2005-1.pdf#page=27. Consultado el 11 de febrero de 2012.
- Bates, D.** 2008. lmer for SAS PROC MIXED Users. Disponible en <http://cran.r-project.org/web/packages/SASmixed/vignettes/Usinglmer.pdf>. Consultado el 6 de noviembre de 2012.
- Bates, D., Machler, M. y Bolker, B.** . 2012. Fitting linear mixed – effects models using lme4. Journal of Statistical Software (Estados Unidos). Disponible en http://listengine.tuxfamily.org/lists.tuxfamily.org/eigen/2011/06/pdfKU_S0z6LjT.pdf. Consultado el 10 de octubre de 2012.
- Burnham, K. y Anderson, D.** 2004. Multimodel Inference. Understanding AIC and BIC in Model Selection. SOCIOLOGICAL METHODS & RESEARCH. 33 (2): 261 – 304. Disponible en

http://faculty.washington.edu/skalski/classes/QERM597/papers_xtra/Burnham%20and%20Anderson.pdf. Consultado el 12 de abril de 2013.

Caballero, F. 2011. Selección de modelos mediante criterios de información en análisis factorial. Aspectos teóricos y computacionales. Tesis Doctoral. Universidad de Granada. Granada, España. 171 p. Disponible en <http://digibug.ugr.es/bitstream/10481/19428/1/19964808.pdf>. Consultado el 12 de abril de 2013.

Cadena, J. y Castillo, A. 2000. Una comparación de SAS y HARVEY en la estimación de componentes de varianza en modelos mixtos. *Agrociencia (México)*. 34: 57 – 68. Disponible en <http://www.colpos.mx/agrocien/Bimestral/2000/ene-feb/art-7.pdf>. Consultado el 10 de abril de 2013.

Cayuela, L. 2010. Modelos lineales mixtos en R. Disponible en <http://158.49.96.73:8080/documenta/bitstream/00000001/26/1/5-Modelos%20lineales%20mixtos%20en%20R.pdf>. Consultado el 26 de noviembre de 2012.

de Graft, H. 2010. Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*. 2 (1): 001 – 006. Disponible en <http://www.academicjournals.org/jdae/PDF/Pdf2010/Jan/Acquah.pdf>. Consultado el 20 de marzo de 2013.

Di Rienzo, J., Macchiavelli, R. y Casanoves, F. 2011. Modelos Lineales Mixtos: Aplicaciones en Infostat. Primera edición. 248 p.

- Dickey, D.** 2008. PROC MIXED: Underlying Ideas with Examples. Disponible en <http://www2.sas.com/proceedings/forum2008/374-2008.pdf>. Consultado el 8 de octubre de 2012.
- Gil, J.** 2001. Comparación de los procedimientos GLM y MIXED del SAS[®] para analizar diseños de parcelas divididas con bloques al azar. *Zootecnia Tropical (Venezuela)*. 19 (1): 43 – 58. Disponible en <http://www.bioline.org.br/request?zt01004>. Consultado el 21 de diciembre de 2012.
- Goodnight, J. H.** 1978. SAS[®] Technical Report 105: Computing MIVQUEO Estimates of Variance Components. SAS Institute Inc. North Carolina, Estados Unidos. 14 p. Disponible en http://support.sas.com/documentation/onlinedoc/v82/techreport_r105.pdf. Consultado el 21 de abril de 2013.
- Hamer, R. y Simpson, P. M.** 2002. Mixed – up Mixed Models: Things That Look Like They Should Work But Don't, and Things That Look Like They Shouldn't Work But Do. Disponible en <http://www2.sas.com/proceedings/sugi25/25/aa/25p020.pdf>. Consultado el 5 de mayo de 2012.
- Hector, A., von Felten, S. y Schmid, B.** 2010a. Analysis of variance with unbalanced data: an update for ecology & evolution. *Journal of Animal Ecology (Reino Unido)*. 79: 308 – 316. Disponible en <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2009.01634.x/abstract>. Consultado el 8 de junio de 2012.
- Hector, A., von Felten, S. y Schmid, B.** 2010b. SUPPLEMENTARY MATERIAL FOR: Analysis of variance with unbalanced data: an update for ecology & evolution. Disponible en

http://www.oikostat.ch/pdfs/hector%20et%20al.%202010JAnimEcol_supplement.pdf.

Consultado el 8 de junio de 2012.

Hu, S. 2012. Akaike Information Criterion. Disponible en http://www4.ncsu.edu/~shu3/Presentation/AIC_2012.pdf. Consultado el 8 de marzo de 2013.

IBM. 2002. Suma de cuadrados. Disponible en http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fglm_general_factorial_sum_of_squares.htm. Consultado el 1 de febrero de 2013.

Junhuan, P., Yun, S., Shuhui, L. y Honglei, Y. 2011. MINQUE of Variance – Covariance Components in Linear Gauss – Markov Models. Journal of Surveying Engineering (Estados Unidos). 137 (4): 129 – 139. Disponible en http://www.cugb.edu.cn/upload/20600/papers_upload/news_20124162056.pdf. Consultado el 22 de abril de 2013.

Kaps, M. y Lamberson, W. 2009. Biostatistics for animal science: an introductory text. Segunda edición. CABI. Oxfordshire. Reino Unido. 504 p.

Kiernan, K., Tao, J. y Gibbs, P. 2012. Tips and Strategies for Mixed Modeling with SAS/STAT © Procedures. Disponible en <http://es.scribd.com/doc/98018739/Tips-and-Strategies-for-Mixed-Modeling-With-SAS-STAT-Procedure>. Consultado el 12 de diciembre de 2012.

- Knaub, J.** 2009. Properties of Weighted Least Squares Regression for Cutoff Sampling in Establishment Surveys. Disponible en <http://interstat.statjournals.net/YEAR/2009/articles/0912003.pdf>. Consultado el 12 de abril de 2013.
- Kuehl, R.** 2001. Diseño de experimentos: principios estadísticos de diseño y análisis de investigación. Segunda edición. Thomson Learning. México, D. F., México. 666 p.
- Langsrud, O.** 2003. ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and computing (Países Bajos)*. 13: 163 – 167. Disponible en <http://wenku.baidu.com/view/4e8edd63caaed3383c4d30f.html>. Consultado el 10 de junio de 2012.
- Liddle, A.** 2008. Information criteria for astrophysical model selection. Disponible en <http://arxiv.org/abs/astro-ph/0701113>. Consultado el 12 de marzo de 2013.
- Littell, R., Freund, R. y Spector, P.** 1991. SAS System for Linear Models. Tercera edición. SAS Institute Inc. North Carolina. Estados Unidos. 329 p.
- Littell, R., Milliken, G., Stroup, W., Wolfinger, R. y Schabenberger, O.** 1996. SAS for Mixed Models: Second edition. SAS Institute Inc. North Carolina. Estados Unidos. 840 p.
- Little, T y Hills, F.** 2002. Métodos estadísticos para la investigación en la agricultura. Segunda edición. Editorial Trillas. México, D. F., México. 270 p.
- Logan, M.** 2010. Biostatistical Design and Analysis Using R: a practical guide. Primera edición. Wiley – Blackwell. Chichester, Reino Unido. 546 p.

- Macnaughton, D.** 1998. Which Sums of Squares Are Best In unbalanced Analysis of Variance. Disponible en <http://www.matstat.com/ss/easleaa0.pdf>. Consultado el 7 de mayo de 2012.
- Miranda, H.** 2011a. Metodología de investigación: Selección de métodos estadísticos aplicados frecuentemente en investigación empírica cuantitativa. Guía del curso Análisis de datos e interpretación de resultados. Facultad de Ciencias Agropecuarias y Forestales. Universidad de La Frontera.
- Miranda, H.** 2011b. Elementos de Diseño y Análisis de Experimentos aplicados a investigación. Guía del curso Análisis de datos e interpretación de resultados. Facultad de Ciencias Agropecuarias y Forestales. Universidad de La Frontera.
- Montgomery, D.** 1991. Diseño y Análisis de Experimentos. Grupo Editorial Iberoamérica. México D. F., México. 589 p.
- Morales, E.** 2005. Diseño experimental a través del análisis de varianza y modelo de regresión lineal: una guía práctica para entender la ciencia de la estadística. Primera edición. Valdivia, Chile. 252 p.
- Motulsky, H. y Christopoulos, A.** 2003. Fitting Models to Biological Data using Linear and Nonlinear Regression: A practical guide to curve fitting. GraphPad Software, Inc. San Diego, CA. Estados Unidos. 351 p. Disponible en <http://www.graphpad.com/manuals/prism4/regressionbook.pdf>. Consultado el 7 de marzo de 2013.
- Piepho, H. P., Buchse, A. y Emrich, K.** 2003. A Hitchhiker's Guide to Mixed Models for Randomized Experiments. J. Agronomy & Crop Science (Alemania). 189: 310 – 322. Disponible en <http://onlinelibrary.wiley.com/doi/10.1046/j.1439-037X.2003.00049.x/abstract;jsessionid=445C695FF5E64F6FC07D19C9D4B9157E.d>

02t01?deniedAccessCustomisedMessage=&userIsAuthenticated=false. Consultado el 8 de julio de 2012.

Posada, D. y Buckley, T. 2004. Model Selection and Model Averaging in Phylogenetics: Advantage of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Test. *SYSTEMATIC BIOLOGY* (Reino Unido). 53 (5): 793 – 808. Disponible en http://xyala2.bio.ed.ac.uk/teaching/tutorials/phylogenetics/Bayesian_Workshop/PDFs/Posada%20and%20Buckley%20Syst%20Biol%202004.pdf. Consultado el 12 de marzo de 2012.

Posada, S. y Rosero, R. 2007. Comparación de modelos matemáticos: una aplicación en la evaluación de alimentos para animales. *Revista Colombiana de Ciencias Pecuarias* (Colombia). 20: 141 – 148. Disponible en <http://rccp.udea.edu.co/index.php/ojs/article/viewFile/270/267>. Consultado el 20 de marzo de 2013.

Quinn, G. y Keough, M. 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press. Primera Edición. Nueva York, Estados Unidos. 537 p. Disponible en http://www.agri.ankara.edu.tr/fcrops/1289_BIYOLOGLAR_ICIN_DENEME_DESENI.pdf. Consultado el 22 de abril de 2013.

Ramírez, G. y Ramírez, B. 2006. Colinealidad y mínimos cuadrados ponderados. *Revista Venezolana de Análisis de Coyuntura* (Venezuela). 12 (1): 283 – 296. Disponible en <http://www.redalyc.org/articulo.oa?id=36412113>. Consultado el 25 de marzo de 2013.

Ramroop, S. 2002. An approach to estimating the variance components to unbalanced cluster sampled survey data and simulated data. Tesis Master of Science in Statistics. University of South Africa. 106 p. Disponible en

<http://uir.unisa.ac.za/bitstream/handle/10500/762/dissertation.pdf?sequence=1>.

Consultado el 22 de abril de 2013.

Rasch, D. y Masata, O. 2006. Methods of variance component estimation. Czech journal of Animal Science (República Checa). 51 (6): 227 – 235. Disponible en <http://agriculturejournals.cz/publicFiles/52286.pdf>. Consultado el 3 de marzo de 2013.

Restrepo, L. 2007. Tipos de sumas de cuadrados en el análisis de varianza. Revista Colombiana de Ciencias Pecuarias (Colombia). 20: 209 – 105. Disponible en http://www.scielo.org.co/scielo.php?pid=S0120-06902007000200013&script=sci_arttext. Consultado el 21 de mayo de 2012.

Rodríguez, J. 2000. Métodos de investigación pecuaria. Editorial Trillas. México, D. F., México. 208 p.

Sanni, S. y Ukaegbu, E. 2012. On Three – Way Unbalance Nested Análisis of Variante. Journal of Mathematics and Statistics (Australia). 8 (1): 1 – 14. Disponible en <http://thescipub.com/abstract/10.3844/jmssp.2012.1.14>. Consultado el 12 de septiembre de 2012.

SAS Institute Inc. 2009. The Four Types of Estimable Functions. Disponible en http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#introglmest_toc.htm. Consultado el 2 de diciembre de 2012.

SAS Institute Inc. 2010a. JMP ® 9: Modeling and Multivariate Methods. Cary, NC: SAS Institute Inc. Disponible en http://support.sas.com/documentation/onlinedoc/jmp/902/Modeling_Multivariate.pdf. Consultado el 12 de enero de 2012.

SAS Institute Inc. 2010b. SAS/STAT ® 9.22 User's Guide: The GLM procedure (Book Excerpt). Disponible en <http://support.sas.com/documentation/cdl/en/statugglm/63679/PDF/default/statugglm.pdf>. Consultado el 6 de mayo de 2012.

SAS Institute Inc. 2010b. SAS/STAT ® 9.22 User's Guide: Introduction to Mixed Modeling Procedures (Book Excerpt). Disponible en <http://support.sas.com/documentation/cdl/en/statugmixedmodel/63718/PDF/default/statugmixedmodel.pdf>. Consultado el 10 de mayo de 2012.

Searle, S. 1987. Linear Models for Unbalanced Data. JOHN WILEY & SONS. Ithaca. New York. 533 p.

Shaw, R. y Mitchell – Olds, T.1993. Anova for unbalanced Data: An Overview. Ecology (Estados Unidos). 74 (6): 1638 – 1645. Disponible en <http://www.jstor.org/discover/10.2307/1939922?uid=2&uid=4&sid=21102263711557>. Consultado el 7 de octubre de 2012.

Solomon, P. J. 2005. Variance Components. Enciclopedia of Biostatistics (Reino Unido). 8: 5685 – 5697. Disponible en http://www.maths.adelaide.edu.au/patty.solomon/EoBII_varcomp.pdf. Consultado el 21 de abril de 2013.

Sosa, W. 2009. Heteroskedasticity and Weighted Least Squares. Disponible en <http://www.econ.uiuc.edu/~wsosa/econ471/GLSHeteroskedasticity.pdf>. Consultado el 26 de abril de 2013.

Spiegel, M., Schiller, J. y Srinivasan, R. 2010. Probabilidad y estadística. Tercera edición. McGraw Hill. México, D. F., México. 425 p.

Spilke, J., Piepho, H. P. y Hu, X. 2005. Analysis of Unbalanced Data by Mixed Linear Models Using the MIXED Procedure of the SAS System. *J. Agronomy & Crop Science* (Alemania). 191: 47 – 54. Disponible en <http://onlinelibrary.wiley.com/doi/10.1111/j.1439-037X.2004.00120.x/abstract>. Consultado el 24 de mayo de 2012.

Wackerly, D., Mendenhall, W. y Scheaffer, R. 2002. Estadística matemática con aplicaciones. Sexta Edición. Thomson Learning. México, D. F., México. 854 p.

Walpole, R., Myers, R., Myers, S. y Ye, K. 2007. Probabilidad y estadística para ingeniería y ciencias. Octava edición. Pearson Educación. México. 840 p.

9. ANEXOS

Cuadro 3. Estimaciones sumas de cuadrados en PROC MIXED y R.

ANOVA SS tipo III de SAS PROC MIXED					ANOVA SS tipo III de R				
Type 3 Tests of Fixed Effects					Analysis of Variance Table				
Effect	Num DF	Den DF	F Value	Pr > F	Df	Sum Sq	Mean Sq	F value	
trt	3	16	1.00	0.4189	4	6.4957	1.62392	31.0372	
iwt	1	16	9.73	0.0066	1	0.5601	0.56005	10.7040	
iwt*trt	3	16	0.95	0.4420	3	0.1484	0.04947	0.9455	

Cuadro 4. Sintaxis en JMP y R.

Sintaxis de JMP	Sintaxis de R
<pre>Fit Model(Y(:adg), Effects(:blk & Random, :trt, :iwt, :trt * :iwt), Personality(Standard Least Squares), Method(REML),);</pre>	<pre>lmer(adg ~ (trt - 1)*iwt + (1 blk), Datos) anova(lmer(adg ~ (trt - 1)*iwt + (1 blk), Datos), type="marginal") pf(c(31.0372), df1=3, df2=16, lower.tail=FALSE) pf(c(10.7040), df1=1, df2=16, lower.tail=FALSE) pf(c(0.9455), df1=3, df2=16, lower.tail=FALSE) lsmeans(lmer(adg ~ (trt - 1)*iwt + (1 blk), Datos), pairwise ~ trt)</pre>

Cuadro 5. Estimaciones de componentes de varianza en JMP y R.

Estimaciones de JMP				Estimaciones de R			
Random Effect	Var Ratio	Component	Var	Random effects:			
blk	4,4165761	0,2310838		Groups	Name	Variance	Std.Dev.
Residual		0,0523219		blk	(Intercept)	0.231084	0.48071
				Residual		0.052322	0.22874

Cuadro 6. Estimaciones sumas de cuadrados en JMP y R.

ANOVA SS tipo III de JMP						ANOVA SS tipo III de R					
Source	Nparm	DF	DFDen	F Ratio	Prob > F	Analysis of Variance Table					
trt	3	3	15,71	10,1855	0,0006*		Df	Sum Sq	Mean Sq	F value	
iwt	1	1	16,37	9,5084	0,0070*	trt	4	6.4957	1.62392	31.0372	
trt*iwt	3	3	16,3	0,9424	0,4430	iwt	1	0.5601	0.56005	10.7040	
						trt:iwt	3	0.1484	0.04947	0.9455	

Cuadro 7. Criterio de Información de Akaike y Criterio de Información Bayesiano en PROC MIXED, JMP y R.

AIC y BIC de SAS PROC MIXED	
AIC (smaller is better)	69.1
AICC (smaller is better)	69.7
BIC (smaller is better)	69.3
AIC y BIC de JMP	
AICc	BIC
101,6687	105,0086

AIC y BIC de R

AIC	BIC
85.12	99.46

Anexo 2. Diseño de un factor fijo y un factor aleatorio.

Cuadro 8. Descripción del problema y sintaxis en PROC MIXED y R.

<p>La variable de respuesta es una calificación. El efecto de las máquinas es fijo y el de los operarios aleatorio, siendo la interacción entre ambos aleatoria.</p>	
<p>Sintaxis de SAS PROC MIXED</p> <pre>proc mixed data=machine method=type3; class machine person; model rating = machine; random person machine*person; lsmeans machine; run;</pre>	<p>Sintaxis de R</p> <pre>lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos) anova(lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos), type="marginal") pf(c(19.966), df1=2, df2=10, lower.tail=FALSE) lsmeans(lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos), pairwise ~ machine)</pre>

Cuadro 9. Estimaciones de componentes de varianza en PROC MIXED y R.

Estimaciones de SAS PROC MIXED		Estimaciones de R			
Cov Parm	Estimate	Random effects:			
person	24.2571	Groups	Name	Variance	Std.Dev.
machine*person	17.0791	employee:machine	(Intercept)	14.23401	3.7728
Residual	0.8726	employee	(Intercept)	22.45579	4.7388
		Residual		0.87087	0.9332

Cuadro 10. Estimaciones sumas de cuadrados en PROC MIXED y R.

ANOVA SS tipo III de SAS PROC MIXED					ANOVA SS tipo III de R					
Effect	Num DF	Den DF	F Value	Pr > F	Analysis of Variance Table					
					Df	Sum Sq	Mean Sq	F value		
machine	2	10	16.72	0.0006	machine	2	34.776	17.388	19.966	

Cuadro 11. Sintaxis en JMP y R.

Sintaxis de JMP	Sintaxis de R
<pre>Fit Model(Y(:overall_rating), Effects(:machine, :employee & Random, :machine * :employee & Random), Personality(Standard Least Squares), Method(REML),);</pre>	<pre>lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos) anova(lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos), type="marginal")</pre>

	<pre> pf(c(19.966), df1=2, df2=10, lower.tail=FALSE) lsmeans(lmer(overall_rating ~ machine + (1 employee) + (1 employee:machine), data=Datos), pairwise ~ machine) </pre>
--	---

Cuadro 12. Estimaciones de componentes de varianza en JMP y R.

Estimaciones de JMP			Estimaciones de R			
		Var	Random effects:			
Random Effect	Var Ratio	Component	Groups	Name	Variance	Std.Dev.
employee	25,785476	22,455765	employee:machine	(Intercept)	14.23401	3.7728
machine*employee	16,344588	14,233991	employee	(Intercept)	22.45579	4.7388
Residual		0,8708687	Residual		0.87087	0.9332

Cuadro 13. Estimaciones sumas de cuadrados en JMP y R.

ANOVA SS tipo III de JMP						ANOVA SS tipo III de R				
Source	Nparm	DF	DFDen	F Ratio	Prob > F	Analysis of Variance Table				
machine	2	2	10,11	19,9639	0,0003*	Df	Sum Sq	Mean Sq	F value	value
						machine	2	34.776	17.388	19.966

Cuadro 14. Criterio de Información de Akaike y Criterio de Información Bayesiano en PROC MIXED, JMP y R.

AIC y BIC de SAS PROC MIXED	
AIC (smaller is better)	188.0
AICC (smaller is better)	188.7
BIC (smaller is better)	187.4

AIC y BIC de JMP	
AICc	BIC
198,339	206,7739

AIC y BIC de R	
AIC	BIC
193.9	204.6

Anexo 3. Diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.

Cuadro 15. Descripción del problema y sintaxis en PROC MIXED y R.

<p>Se comparan dos tratamientos en cada una de siete granjas, donde los tratamientos son considerados fijos, mientras las granjas son consideradas efectos aleatorios.</p>	
<p>Sintaxis de SAS PROC MIXED</p> <pre> proc mixed data=farm; class farm rep trt; model y=trt; random rep farm rep(farm) farm*trt; lsmeans trt; run; </pre>	<p>Sintaxis de R</p> <pre> lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos) anova(lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos), type="marginal") pf(c(2.9637), df1=1, df2=6, lower.tail=FALSE) lsmeans(lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos), pairwise ~ trt) </pre>

Cuadro 16. Estimaciones de componentes de varianza en PROC MIXED y R.

Estimaciones de SAS PROC MIXED		Estimaciones de R			
Cov Parm	Estimate	Random effects:			
rep	1.5434	Groups	Name	Variance	Std.Dev.
farm	7.3813	farm:trt	(Intercept)	43.3505	6.5841
rep(farm)	1.4623	farm:rep	(Intercept)	1.4622	1.2092
farm*trt	43.3507	farm	(Intercept)	7.3818	2.7169
Residual	15.2256	rep	(Intercept)	1.5434	1.2423
		Residual		15.2256	3.9020

Cuadro 17. Estimaciones sumas de cuadrados en PROC MIXED y R.

ANOVA SS tipo III de SAS PROC MIXED					ANOVA SS tipo III de R			
Type 3 Tests of Fixed Effects					Analysis of Variance Table			
Effect	Num DF	Den DF	F Value	Pr > F	Df	Sum Sq	Mean Sq	F value
trt	1	6	2.96	0.1359	trt 1	45.124	45.124	2.9637

Cuadro 18. Sintaxis en JMP y R.

Sintaxis de JMP	Sintaxis de R
<pre>Fit Model(Y(:y), Effects(:farm & Random, :trt, :trt * :farm & Random, :rep[:farm] & Random, :rep & Random), Personality(Standard Least Squares), Method(REML),);</pre>	<pre>lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos) anova(lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos), type="marginal") pf(c(2.9637), df1=1, df2=6, lower.tail=FALSE) lsmeans(lmer(Y ~ trt+(1 rep/farm)+(1 farm)+(1 farm:trt), data=Datos), pairwise ~ trt)</pre>

Cuadro 19. Estimaciones de componentes de varianza en JMP y R.

Estimaciones de JMP			Estimaciones de R			
Random Effect	Var Ratio	Var Component	Random effects:			
farm	0,4847947	7,3812991	Groups	Name	Variance	Std.Dev.
trt*farm	2,8472194	43,350676	farm:trt	(Intercept)	43.3505	6.5841
rep[farm]	0,0960417	1,4622946	farm:rep	(Intercept)	1.4622	1.2092
rep	0,10137	1,5434202	farm	(Intercept)	7.3818	2.7169
Residual		15,225618	rep	(Intercept)	1.5434	1.2423
			Residual		15.2256	3.9020

Cuadro 20. Estimaciones sumas de cuadrados en JMP y R.

ANOVA SS tipo III de JMP						ANOVA SS tipo III de R				
Source	Nparm	DF	DFDen	F Ratio	Prob > F	Analysis of Variance Table				
trt	1	1	5,563	2,9235	0,1420	Df	Sum Sq	Mean Sq	F value	
						trt	1	45.124	45.124	2.9637

Cuadro 21. Criterio de Información de Akaike y Criterio de Información Bayesiano en PROC MIXED, JMP y R.

AIC y BIC de SAS PROC MIXED	
AIC (smaller is better)	167.2
AICC (smaller is better)	170.7
BIC (smaller is better)	162.7
AIC y BIC de JMP	
AICc	BIC
179,1874	181,1313
AIC y BIC de R	
AIC	BIC
171.2	179.7

Anexo 4. Diseño de parcelas divididas.

Cuadro 22. Descripción del problema y sintaxis en PROC MIXED y R.

<p>A es el factor asignado a la parcela principal y corresponde a tres niveles de riego, y el factor B, asignado a las subparcelas, representa dos variedades de trigo.</p>	
<p>Sintaxis de SAS PROC MIXED</p> <pre>proc mixed; class A B Block; model Y = A B A*B; random Block A*Block; lsmeans A B A*B; run;</pre>	<p>Sintaxis de R</p> <pre>lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos) anova(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos), type="marginal") pf(c(4.3987), df1=2, df2=6, lower.tail=FALSE) pf(c(17.2661), df1=1, df2=8, lower.tail=FALSE) pf(c(4.1317), df1=2, df2=8, lower.tail=FALSE) lsmeans(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos), list(poly ~ Riego, pairwise ~ Variedad)) lsmeans(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos),</pre>

	list(poly ~ Riego, pairwise ~ Riego:Variedad))
--	--

Cuadro 23. Estimaciones de componentes de varianza en PROC MIXED y R.

Estimaciones de SAS PROC MIXED		Estimaciones de R	
Cov Parm	Estimate	Random effects:	
Block	50.9788	Groups	Name
A*Block	12.9034	Riego:block	(Intercept)
Residual	6.6968	block	(Intercept)
		Residual	
			Variance Std.Dev.
			12.9032 3.5921
			50.9938 7.1410
			6.6963 2.5877

Cuadro 24. Estimaciones sumas de cuadrados en PROC MIXED y R.

ANOVA SS tipo III de SAS PROC MIXED					ANOVA SS tipo III de R				
Effect	Num DF	Den DF	F Value	Pr > F	Analysis of Variance Table				
A	2	6	4.53	0.0631		Df	Sum Sq	Mean Sq	F value
B	1	8	17.66	0.0030	Riego	2	58.910	29.455	4.3987
A*B	2	8	4.13	0.0586	Variedad	1	115.620	115.620	17.2661
					Riego:Variedad	2	55.334	27.667	4.1317

Cuadro 25. Sintaxis en JMP y R.

Sintaxis de JMP	Sintaxis de R
<pre>Fit Model(Y(:Y), Effects(:Riego, :Variedad, :Riego * :Variedad, :block & Random, :block * :Riego & Random), Personality(Standard Least Squares), Method(REML),);</pre>	<pre>lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos) anova(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos), type="marginal") pf(c(4.3987), df1=2, df2=6, lower.tail=FALSE) pf(c(17.2661), df1=1, df2=8, lower.tail=FALSE) pf(c(4.1317), df1=2, df2=8, lower.tail=FALSE) lsmeans(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos), list(poly ~ Riego, pairwise ~ Variedad)) lsmeans(lmer(Y ~ Riego + Variedad + Riego:Variedad + (1 block) + (1 Riego:block), data = Datos), list(poly ~ Riego, pairwise ~ Riego:Variedad))</pre>

Cuadro 26. Estimaciones de componentes de varianza en JMP y R.

Estimaciones de JMP			Estimaciones de R			
Random Effect	Var Ratio	Var Component	Random effects:			
block	7,6151785	50,993788	Groups	Name	Variance	Std.Dev.
block*Riego	1,9269109	12,903241	Riego:block	(Intercept)	12.9032	3.5921
Residual		6,6963352	block	(Intercept)	50.9938	7.1410
			Residual		6.6963	2.5877

Cuadro 27. Estimaciones sumas de cuadrados en JMP y R.

ANOVA SS tipo III de JMP						ANOVA SS tipo III de R				
Source	Nparm	DF	DFDen	F Ratio	Prob > F	Analysis of Variance Table				
Riego	2	2	6,148	4,5326	0,0617		Df	Sum Sq	Mean Sq	F value
Variedad	1	1	8,275	17,4549	0,0029*	Riego	2	58.910	29.455	4.3987
Riego*Variedad	2	2	8,259	4,1272	0,0572	Variedad	1	115.620	115.620	17.2661
						Riego:Variedad	2	55.334	27.667	4.1317

Cuadro 28. Criterio de Información de Akaike y Criterio de Información Bayesiano en PROC MIXED, JMP y R.

AIC y BIC de SAS PROC MIXED	
AIC (smaller is better)	115.3
AICC (smaller is better)	117.1
BIC (smaller is better)	113.4
AIC y BIC de JMP	

AICc	BIC
149,6885	146,0618
AIC y BIC de R	
AIC	BIC
127.3	137.5

Anexo 5. Diseño cross – over.

Cuadro 29. Descripción del problema y sintaxis en PROC MIXED y R.

<p>Se tienen dos tratamientos (A y B) y dos secuencias (AB y BA). Cada uno de los 17 sujetos es sometido a los tratamientos en dos períodos consecutivos.</p>	
<p>Sintaxis de SAS PROC MIXED</p> <pre> PROC MIXED DATA = asthma; CLASS trt sub per seq; MODEL fev = seq trt per; RANDOM sub(seq); LSMEANS seq trt per; RUN; </pre>	<p>Sintaxis de R</p> <pre> lmer(fev ~ seq + trt + per + (1 sub), data=Datos) anova(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), type="marginal") pf(c(2.6347), df1=1, df2=15, lower.tail=FALSE) pf(c(4.9960), df1=1, df2=15, lower.tail=FALSE) pf(c(1.3734), df1=1, df2=15, lower.tail=FALSE) lsmeans(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), list(poly ~ seq, pairwise ~ trt)) lsmeans(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), list(poly ~ seq, pairwise ~ per)) </pre>

Cuadro 30. Estimaciones de componentes de varianza en PROC MIXED y R.

Estimaciones de SAS PROC MIXED		Estimaciones de R	
Cov Parm	Estimate	Random effects:	
subject(seq)	0.3618	Groups	Name
Residual	0.1192	sub	(Intercept)
			Variance
		Residual	Std.Dev.
			0.36178
			0.60148
			0.11921
			0.34527

Cuadro 31. Estimaciones sumas de cuadrados en PROC MIXED y R.

ANOVA SS tipo III de SAS PROC MIXED					ANOVA SS tipo III de R				
Effect	Num DF	Den DF	F Value	Pr > F	Analysis of Variance Table				
					Df	Sum Sq	Mean Sq	F value	
seq	1	15	2.64	0.1253	seq	1	0.31409	0.31409	2.6347
trt	1	15	4.68	0.0472	trt	1	0.59559	0.59559	4.9960
period	1	15	1.37	0.2595	per	1	0.16373	0.16373	1.3734

Cuadro 32. Sintaxis en JMP y R.

Sintaxis de JMP	Sintaxis de R
<pre>Fit Model(Y(:fev), Effects(:seq, :trt, :per, :sub[:seq] & Random), Personality(Standard Least Squares),</pre>	<pre>lmer(fev ~ seq + trt + per + (1 sub), data=Datos) anova(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), type="marginal") pf(c(2.6347), df1=1, df2=15,</pre>

<pre>Method(REML),);</pre>	<pre>lower.tail=FALSE) pf(c(4.9960), df1=1, df2=15, lower.tail=FALSE) pf(c(1.3734), df1=1, df2=15, lower.tail=FALSE) lsmeans(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), list(poly ~ seq, pairwise ~ trt)) lsmeans(lmer(fev ~ seq + trt + per + (1 sub), data=Datos), list(poly ~ seq, pairwise ~ per))</pre>
-------------------------------	--

Cuadro 33. Estimaciones de componentes de varianza en JMP y R.

<p>Estimaciones de JMP</p> <table border="1"> <thead> <tr> <th>Random Effect</th> <th>Var Ratio</th> <th>Var Component</th> </tr> </thead> <tbody> <tr> <td>sub[seq]</td> <td>3,0347202</td> <td>0,3617763</td> </tr> <tr> <td>Residual</td> <td></td> <td>0,1192124</td> </tr> </tbody> </table>	Random Effect	Var Ratio	Var Component	sub[seq]	3,0347202	0,3617763	Residual		0,1192124	<p>Estimaciones de R</p> <pre>Random effects: Groups Name Variance Std.Dev. sub (Intercept) 0.36178 0.60148 Residual 0.11921 0.34527</pre>
Random Effect	Var Ratio	Var Component								
sub[seq]	3,0347202	0,3617763								
Residual		0,1192124								

Cuadro 34. Estimaciones sumas de cuadrados en JMP y R.

<p>ANOVA SS tipo III de JMP</p> <table border="1"> <thead> <tr> <th>Source</th> <th>Nparm</th> <th>DF</th> <th>DFDen</th> <th>F Ratio</th> <th>Prob > F</th> </tr> </thead> <tbody> <tr> <td>seq</td> <td>1</td> <td>1</td> <td>15</td> <td>2,6357</td> <td>0,1253</td> </tr> <tr> <td>trt</td> <td>1</td> <td>1</td> <td>15</td> <td>4,6759</td> <td>0,0472*</td> </tr> <tr> <td>per</td> <td>1</td> <td>1</td> <td>15</td> <td>1,3734</td> <td>0,2595</td> </tr> </tbody> </table>	Source	Nparm	DF	DFDen	F Ratio	Prob > F	seq	1	1	15	2,6357	0,1253	trt	1	1	15	4,6759	0,0472*	per	1	1	15	1,3734	0,2595	<p>ANOVA SS tipo III de R</p> <pre>Analysis of Variance Table Df Sum Sq Mean Sq F value seq 1 0.31409 0.31409 2.6347 trt 1 0.59559 0.59559 4.9960 per 1 0.16373 0.16373 1.3734</pre>
Source	Nparm	DF	DFDen	F Ratio	Prob > F																				
seq	1	1	15	2,6357	0,1253																				
trt	1	1	15	4,6759	0,0472*																				
per	1	1	15	1,3734	0,2595																				

Cuadro 35. Criterio de Información de Akaike y Criterio de Información Bayesiano en PROC MIXED, JMP y R.

AIC y BIC de SAS PROC MIXED	
AIC (smaller is better)	64.6
AICC (smaller is better)	65.1
BIC (smaller is better)	66.3
AIC y BIC de JMP	
AICc	BIC
79,8772	85,92425
AIC y BIC de R	
AIC	BIC
72.61	81.77

Anexo 6. Tablas de datos para los ejemplos seleccionados.

Cuadro 36. Datos correspondientes al diseño de un factor con bloque y covarianza.

Nº	Bloque	Tratamiento	GPD	PI
1	B_1	T_0	1,03	338
2	B_1	T_10	1,54	477
3	B_1	T_20	1,82	444
4	B_1	T_30	1,86	370
5	B_2	T_0	1,31	403
6	B_2	T_10	2,16	451
7	B_2	T_20	2,13	450
8	B_2	T_30	2,23	393
9	B_3	T_0	1,59	394
10	B_3	T_10	2,53	499
11	B_3	T_20	2,33	482
12	B_3	T_30	1,8	317
13	B_4	T_0	2,09	499
14	B_4	T_10	2,2	411
15	B_4	T_20	2,21	391
16	B_4	T_30	2,82	396
17	B_5	T_0	1,66	371
18	B_5	T_10	2,3	418
19	B_5	T_20	2,65	486
20	B_5	T_30	2,18	333
21	B_6	T_0	1,42	395
22	B_6	T_10	1,93	325
23	B_6	T_20	1,58	316
24	B_6	T_30	1,49	311
25	B_7	T_0	1,41	414
26	B_7	T_10	1,65	313
27	B_7	T_20	1,08	309

28	B_7	T_30	1,34	323
29	B_8	T_0	0,18	315
30	B_8	T_10	0,64	376
31	B_8	T_20	0,76	308
32	B_8	T_30	0,7	439

Cuadro 37. Datos correspondientes al diseño de un factor fijo y un factor aleatorio.

Nº	Maquina	Empleado	Calificación
1	M_1	E_1	52
2	M_1	E_1	
3	M_1	E_1	
4	M_1	E_2	51,8
5	M_1	E_2	52,8
6	M_1	E_2	
7	M_1	E_3	60
8	M_1	E_3	
9	M_1	E_3	
10	M_1	E_4	51,1
11	M_1	E_4	52,3
12	M_1	E_4	
13	M_1	E_5	50,9
14	M_1	E_5	51,8
15	M_1	E_5	51,4
16	M_1	E_6	46,4
17	M_1	E_6	44,8
18	M_1	E_6	49,2
19	M2	E_1	64
20	M_2	E_1	
21	M_2	E_1	
22	M_2	E_2	59,7
23	M_2	E_2	60
24	M_2	E_2	59
25	M_2	E_3	68,6
26	M_2	E_3	65,8
27	M_2	E_3	
28	M_2	E_4	63,2
29	M_2	E_4	62,8
30	M_2	E_4	62,2

31	M_2	E_5	64,8
32	M_2	E_5	65
33	M_2	E_5	
34	M_2	E_6	43,7
35	M_2	E_6	44,2
36	M_2	E_6	43
37	M_3	E_1	67,5
38	M_3	E_1	67,2
39	M_3	E_1	66,9
40	M_3	E_2	61,5
41	M_3	E_2	61,7
42	M_3	E_2	62,3
43	M_3	E_3	70,8
44	M_3	E_3	70,6
45	M_3	E_3	71
46	M_3	E_4	64,1
47	M_3	E_4	66,2
48	M_3	E_4	64
49	M_3	E_5	72,1
50	M_3	E_5	72
51	M_3	E_5	71,1
52	M_3	E_6	62
53	M_3	E_6	61,4
54	M_3	E_6	60,5

Cuadro 38. Datos correspondientes al diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.

Nº	Granja	Repetición	Tratamiento	Y
1	G_1	R_1	T_1	49,7
2	G_1	R_1	T_2	56,2
3	G_1	R_2	T_1	41
4	G_1	R_2	T_2	52,8
5	G_1	R_3	T_1	40,8
6	G_1	R_3	T_2	57,4
7	G_2	R_1	T_1	42,9
8	G_2	R_1	T_2	55,9
9	G_2	R_2	T_1	47,5
10	G_2	R_2	T_2	52,6
11	G_3	R_1	T_1	41,8
12	G_3	R_1	T_2	46,9
13	G_3	R_2	T_1	39,2
14	G_3	R_2	T_2	33,3
15	G_4	R_1	T_1	31,5
16	G_4	R_1	T_2	51,6
17	G_4	R_2	T_1	33,5
18	G_4	R_2	T_2	54,5
19	G_5	R_1	T_1	46,8
20	G_5	R_1	T_2	36,6
21	G_6	R_1	T_1	49
22	G_6	R_1	T_2	63,9
23	G_7	R_1	T_1	43,2
24	G_7	R_1	T_2	40,2
25	G_7	R_2	T_1	37,8
26	G_7	R_2	T_2	31,9

Cuadro 39. Datos correspondientes al diseño de parcelas divididas.

Nº	Bloque	Riego	Variedad	Y
1	B_1	R_1	V_1	56
2	B_1	R_1	V_2	41
3	B_1	R_2	V_1	50
4	B_1	R_2	V_2	36
5	B_1	R_3	V_1	39
6	B_1	R_3	V_2	35
7	B_2	R_1	V_1	30
8	B_2	R_1	V_2	25
9	B_2	R_2	V_1	36
10	B_2	R_2	V_2	28
11	B_2	R_3	V_1	33
12	B_2	R_3	V_2	30
13	B_3	R_1	V_1	32
14	B_3	R_1	V_2	24
15	B_3	R_2	V_1	31
16	B_3	R_2	V_2	27
17	B_3	R_3	V_1	15
18	B_3	R_3	V_2	19
19	B_4	R_1	V_1	30
20	B_4	R_1	V_2	25
21	B_4	R_2	V_1	35
22	B_4	R_2	V_2	30
23	B_4	R_3	V_1	17
24	B_4	R_3	V_2	18

Cuadro 40. Datos correspondientes al diseño cross – over.

Nº	Secuencia	Sujeto	Tratamiento	Periodo	VAE
1	S_1	S_1	T_1	P_1	1,28
2	S_1	S_1	T_2	P_2	1,33
3	S_1	S_2	T_1	P_1	1,6
4	S_1	S_2	T_2	P_2	2,21
5	S_1	S_3	T_1	P_1	2,46
6	S_1	S_3	T_2	P_2	2,43
7	S_1	S_4	T_1	P_1	1,41
8	S_1	S_4	T_2	P_2	1,81
9	S_1	S_5	T_1	P_1	1,4
10	S_1	S_5	T_2	P_2	0,85
11	S_1	S_6	T_1	P_1	1,12
12	S_1	S_6	T_2	P_2	1,2
13	S_1	S_7	T_1	P_1	0,9
14	S_1	S_7	T_2	P_2	0,9
15	S_1	S_8	T_1	P_1	2,41
16	S_1	S_8	T_2	P_2	2,79
17	S_1	S_9	T_1	P_1	
18	S_1	S_9	T_2	P_2	
19	S_2	S_10	T_2	P_1	3,06
20	S_2	S_10	T_1	P_2	1,38
21	S_2	S_11	T_2	P_1	2,68
22	S_2	S_11	T_1	P_2	2,1
23	S_2	S_12	T_2	P_1	2,6
24	S_2	S_12	T_1	P_2	2,32
25	S_2	S_13	T_2	P_1	1,48
26	S_2	S_13	T_1	P_2	1,3
27	S_2	S_14	T_2	P_1	2,08
28	S_2	S_14	T_1	P_2	2,34
29	S_2	S_15	T_2	P_1	2,72
30	S_2	S_15	T_1	P_2	2,48

31	S_2	S_16	T_2	P_1	1,94
32	S_2	S_16	T_1	P_2	1,11
33	S_2	S_17	T_2	P_1	3,35
34	S_2	S_17	T_1	P_2	3,23
35	S_2	S_18	T_2	P_1	1,16
36	S_2	S_18	T_1	P_2	1,25

Anexo 7. Medias de mínimos cuadrados.

Cuadro 41. LS Means para el diseño de un factor con bloque y covarianza.

LS Means JMP			LS Means R		LS Means SAS	
Tratamiento			Tratamiento		Tratamiento	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
0	1,3333001	0,18822811	1,333300	0,1882237	1,3333	0,1882
10	1,8869865	0,1953528	1,886987	0,1951599	1,8870	0,1952
20	1,790787	0,18854286	1,790787	0,1885298	1,7908	0,1885
30	1,9183811	0,19962781	1,918380	0,1993459	1,9184	0,1993

Cuadro 42. LS Means para el diseño de un factor fijo y un factor aleatorio.

LS Means JMP			LS Means R		LS Means SAS	
Máquina			Máquina		Máquina	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
M1	52,354	2,4906713	52,35400	2,490493	52,3545	2,6416
M2	60,316445	2,4874448	60,31644	2,487285	60,3197	2,6385
M3	66,272222	2,4826077	66,27222	2,482486	66,2722	2,6340

Cuadro 43. LS Means para el diseño de dos factores, con el factor principal aleatorio y el factor secundario fijo.

LS Means JMP			LS Means R		LS Means SAS	
Tratamiento			Tratamiento		Tratamiento	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
1	42,248110	3,1399337	42,24815	3,063346	42,2481	3,0634
2	48,953937	3,2016616	48,95400	3,094590	48,9539	3,0946

Cuadro 44. LS Means para el diseño de parcelas divididas.

LS Means JMP			LS Means R		LS Means SAS	
Riego			Riego		Riego	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
R_1	31,600238	4,1337127	31,60023	4,129593	31,6002	4,1294
R_2	34,125	4,1001584	34,12500	4,099955	34,1250	4,0997
R_3	25,75	4,1001584	25,75000	4,099955	25,7500	4,0997
Variedad			Variedad		Variedad	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
V_1	32,816825	3,8084822	32,81682	3,806374	32,8168	3,8061
V_2	28,166667	3,7923271	28,16667	3,792107	28,1667	3,7918
Riego*Variedad			Riego*Variedad		Riego*Variedad	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
R_1,V_1	34,450476	4,3305273	34,45046	4,315356	34,4504	4,3151
R_1,V_2	28,75	4,2009929	28,75000	4,200795	28,7500	4,2006
R_2,V_1	38	4,2009929	38,00000	4,200795	38,0000	4,2006
R_2,V_2	30,25	4,2009929	30,25000	4,200795	30,2500	4,2006
R_3,V_1	26	4,2009929	26,00000	4,200795	26,0000	4,2006
R_3,V_2	25,5	4,2009929	25,50000	4,200795	25,5000	4,2006

Cuadro 45. LS Means para el diseño cross – over.

LS Means JMP			LS Means R		LS Means SAS	
Secuencia			Secuencia		Secuencia	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
1	1,63125	0,22950558	1,631250	0,2294989	1,6312	0,2295
2	2,1433333	0,21637994	2,143333	0,2163737	2,1433	0,2164
Tratamiento			Tratamiento		Tratamiento	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
1	1,7590278	0,16849835	1,759028	0,1684941	1,7590	0,1685
2	2,0155556	0,16849835	2,015556	0,1684941	2,0156	0,1685
Período			Período		Período	
Nivel	Least Sq Mean	Std Error	Least Sq Mean	Std Error	Least Sq Mean	Std Error
1	1,9568056	0,16849835	1,956806	0,1684941	1,9568	0,1685
2	1,8177778	0,16849835	1,817778	0,1684941	1,8178	0,1685