JOURNAL OF
BASIC AND APPLIED GENETICS
ISSN: 1852-6233

MÉTODOS ESTADÍSTICOS EN GENÉTICA BÁSICA Y APLICADA: POR QUÉ, CÓMO Y CUÁNTO

Babinec F.J.[1]

[1]Estación Experimental Agropecuaria Anguil, INTA
Facultad de Agronomía, UNLPampa.

babinec.francisco@inta.gov.ar / fbabinec@agro.unlpam.edu.ar

"Un hombre con un martillo ve al mundo como un clavo"
Mark Twain

ABSTRACT

This note is an overview of the statistical treatment of research in basic and applied genetics, starting with a brief mention of the types of studies used and their scope, following with some problems in experiments and observational studies, types of variables analyzed and some commonly used methods, and introducing the Bayesian approach. Finally I mention some of the problems currently under study and methodologies used, and give some recommendations at the end.

Key words: experimental design, statistical analysis, Bayesian methods, non-parametrical methods, mixed models, statistical software.

RESUMEN

Se presenta una revisión general del tratamiento estadístico de las investigaciones en genética básica y aplicada, partiendo de una breve mención a los tipos de estudios posibles y su ámbito de aplicación. Se mencionan luego algunos problemas en experimentos y estudios observacionales, tipos de variables analizadas y algunos métodos de uso común, para luego introducir el enfoque bayesiano. Se mencionan finalmente algunos de los problemas bajo estudio actualmente y algunas metodologías empleadas, para dar al final algunas recomendaciones.

Palabras clave: diseño experimental, análisis estadístico, métodos bayesianos, métodos no paramétricos, modelos mixtos, programas estadísticos.

INTRODUCCIÓN

La Genética y la Estadística han estado estrechamente relacionadas en su desarrollo durante el siglo XX (Nelsen, T.C. 2002). Es más, hasta el advenimiento de la revolución molecular, la Genética tenía una profunda impronta probabilística (Fisher, 1918, 1930; Mather, 1938). Algunos de los mayores desarrollos de la Estadística, por otro lado, se deben a genetistas como Henderson y otros (Pianola, 2002; Littell 2011). En verdad, la Estadística ha penetrado toda la investigación en el siglo pasado y se ha convertido en el estilo de pensamiento científico dominante, para usar la terminología de Crombie (*cf.* Hacking, 2007). Pero ello ha implicado también un uso y abuso de las técnicas estadísticas en muchas áreas, como la agronomía (Nelson y Rawlings, 1983; Maindonald, 1984; Dyke 1997), medicina (Altman, 1982), psicología (Hager, 2000), etc. Los problemas encontrados por éstos y otros autores van de la planificación inadecuada de las experiencias a la elección de métodos de análisis inapropiados y la interpretación incorrecta de los resultados, como resume didácticamente Warren (1986). Como reacción, muchas revistas científicas, como por ejemplo el *British Medical Journal*, han publicado series de artículos dedicados a tratar conceptos básicos de estadística y presentar los métodos de análisis más usados en la respectiva disciplina: ejemplos (Platt, (1997, 1998 a, 1998b); recomendaciones (Altman *et al.*, 1983; Bailar y Mosteller, 1988; Wilkinson, 1999; Gould y Steiner, 2003); revisiones (Onofri *et al.*, 2010; van Putten *et al.*, 2010) o puestas al día presentando nuevas metodologías (Matson *et al.*, 1993; Garrett *et al.*, 2004; Machado y Petrie, 2006). Dados los continuos avances en ambas disciplinas, resulta conveniente revisar las aplicaciones más comunes de la estadística a las investigaciones en genética básica y aplicada. A lo largo del tiempo han aparecido revisiones parciales sobre el tratamiento estadístico de datos genéticos (por ej. Fisher, 1952; Lin *et al.*, 1986; Kearsey y Farquhar, 1998; Balding, 2006; Montana, 2006; Terwilliger y Göring, 2009; Balzarini *et al.*, 2011). Esta es una presentación general del tema, para desarrollar en notas posteriores algunas metodologías desde el punto de vista de los problemas más frecuentes a investigar. Somos conscientes de que sólo es posible cubrir una fracción del tema (Potvin y Travis, 1993); la bibliografía citada es sólo una muestra y refleja el (des)conocimiento del autor sobre la materia.

MÉTODOS ESTADÍSTICOS E INVESTIGACIONES EN GENÉTICA BÁSICA Y APLICADA

El punto de partida

Muchas de las revisiones sobre empleo de la estadística en distintas disciplinas ponen énfasis en las fallas en el análisis e interpretación. Pero el primer paso donde interviene la estadística en una investigación científica debería ser en la planificación cuidadosa de la experiencia a realizar, sea la elección de los tratamientos, si fuera un ensayo, o de las poblaciones a observar, si fuera un estudio. La diferenciación entre estudios observacionales y experimentos es crucial desde el punto de vista de las inferencias que podrán extraerse finalmente de las diferentes experiencias. En los experimentos el investigador controla ciertas variables o condiciones y puede postular relaciones causa-efecto, mientras que en los estudios observacionales puede identificar factores de riesgo o condiciones predisponentes (Scheiner, 1993). En el caso de la genética, que es básicamente el estudio de la transmisión de características a través de las generaciones (Scheiner, 2010), en general mediante la crianza experimental (Bateson, 1908), la condición experimental está dada por la elección de los progenitores y la determinación del sistema de apareamiento y las progenies (generaciones) segregantes, mientras que en los estudios observacionales será la identificación de las poblaciones caso y control. La elección de los progenitores define la población de referencia, tema objeto de amplio debate (Cockerham, 1980). La relación entre el conjunto de individuos estudiados y la población hacia la cual se hará inferencia lleva a la distinción entre modelos de efectos fijos, cuando el conjunto constituye toda la población de referencia, y aleatorios, cuando son una muestra de la misma. En los modelos mixtos hay una combinación de ambos tipos de efectos (Mc Lean *et al.*, 1991). Otros tipos de estudios son la integración de resultados publicados previamente o meta-análisis y el análisis de bases de datos.

La técnica experimental

Se trate de experimentos o estudios, hay cuestiones comunes en su planificación, como la determinación del tamaño de muestra necesario para estimar una proporción o para representar una población (Hanson, 1959) o el número de replicaciones (Geng y Hills, 1978; Gauch y Zobel, 1996). La evaluación imparcial de las progenies segregantes requiere del diseño de experimentos apropiados, que cumplan los requisitos de replicación, aleatorización y control local establecidos por Fisher (1960; ver también White, 1984; St-Pierre, 2007; Eskridge 2009). Desde el advenimiento de los ensayos en bloques se han registrado avances tanto en la generación de nuevos diseños para elevado número de tratamientos (Edmondson, 2005; Eskridge, 2009) como en la incorporación de nuevas técnicas de análisis (Lawson y Cressie, 2000; Piepho *et al.*, 2003, 2004; Payne, 2006; Hua y Spilke, 2011). Algunos de los problemas que pueden aparecer en las estimaciones de parámetros genéticos en plantas vinculados con el diseño y la técnica experimentales han sido revisados por Moll y Robinson (1967), Dudley y Moll (1969) y Holland *et al.* (2003), entre otros. Un aspecto que no siempre es tenido en cuenta es la vinculación entre el sitio experimental u observacional y el universo al que se desea inferir (Templeman, 2009). Los estudios a nivel molecular plantean nuevas dificultades tanto en diseño como en análisis (Churchill, 2002; Quackenbush, 2002; Rosa *et al.*, 2005; para citar algunos).

Variables y análisis

El objetivo de una experiencia es la recolección de datos sobre los individuos estudiados y su interpretación. Según respondan a características preexistentes u observadas durante la marcha de la experiencia o al final de la misma, hablamos de variables clasificatorias, predictivas, explicativas o independientes y/o de covariables, y de variables respuesta o dependientes, aunque esta distinción es difusa en muchas oportunidades. En ocasiones, las variables registradas son sólo una aproximación al verdadero sujeto de estudio, o una construcción mental, como el rendimiento. Las variables, sean dependientes o independientes, pueden tomar sólo dos valores (binarias) o varios (categóricas, nominales) siguiendo a veces un orden (ordinales) o pueden tomar cualquier valor en una escala numérica (continuas). Esta clasificación sigue en gran parte la terminología introducida por Stevens (1946), quien tuvo en cuenta el efecto de distintas transformaciones sobre el monto de información que contenía cada variable. Una transformación es "permisible" o "admisible" en la medida que preserva dicho monto. Posteriormente, la extensión de estos conceptos a los métodos de análisis (Stevens, 1951, citado por Velleman y Wilkinson, 1993) dio lugar a la presentación en varios textos de tablas de métodos "admisibles", que son seguidas a veces como un canon que estipula los únicos modos posibles de analizar los datos obtenidos, en función de la escala en la que están medidas las variables de interés, sin tener en cuenta el continuo desarrollo de nuevos métodos estadísticos, y la revisión de otros ya establecidos (Cohen, 2001). Un caso especial es el de los datos de supervivencia o confiabilidad (censurados a la derecha) que han dado lugar al desarrollo de métodos específicos (Cox, 1972). Un ejemplo de lo complejo que resulta discernir el tipo de variable medida está dado por la resistencia a enfermedades, que si bien se registra siguiendo escalas más o menos aceptadas, presenta en muchos casos una suerte de *continuum* en la magnitud de las lesiones observadas.

El análisis estadístico

El primer paso del análisis, como se enseña en los cursos de estadística aplicada, es realizar un análisis exploratorio a fin de detectar valores extraños debidos a observaciones erróneas u otros accidentes experimentales (Platt, 1998b). A las herramientas gráficas disponibles (histogramas, gráficos de caja, de tallo y hoja, de dispersión, etc.) pueden sumarse algunas técnicas analíticas como el análisis de correlación, y ciertos métodos multivariados como los análisis de componentes principales y de correspondencia, que permiten la reducción de dimensiones, algo necesario cuando hay muchas variables respuesta medidas.

El análisis estadístico puede verse como la búsqueda de un modelo, o sea, el ajuste de los datos obtenidos a alguna distribución teórica. Dicho de

otro modo, se busca explicar la variabilidad presente en la variable respuesta, separándola en partes que pueden ser explicadas de otras que no pueden serlo, lo que es la base del análisis de varianza (ANOVA). La estadística aplicada se basa en la estimación de parámetros y en la prueba de hipótesis, debidas a R. Fisher y a J. Neyman y E. Pearson, respectivamente; ambos casos requieren cumplimentar algunos supuestos, en particular sobre la distribución de la parte no explicada por el modelo propuesto. Estos supuestos, que en sí son condiciones indemostrables (Mead, 1990), se refieren tanto a la independencia de las observaciones como a la distribución de los valores observados y/o de los residuales del modelo, o sea de la fracción no explicada por el mismo. Algunas técnicas, como el ANOVA, requieren una distribución normal (gaussiana) de los residuales, a diferencia de otras a las que se denomina no-paramétricas o de distribución libre. El estudio de los residuales puede llevarnos a transformar la escala de la variable respuesta, de la explicativa, o de ambas, para ajustar una distribución normal de los mismos (Fernández, 1992), o a la opción de métodos no paramétricos (Potvin y Roff, 1993) o al empleo de modelos mixtos (Paterson y Lello, 2003; Piepho, 2003; 2004) o generalizados (Paterson y Lello, 2003; Bolker *et al.*, 2008), en lugar del tradicional ANOVA.

La Tabla 1, adaptada de Motulsky (1995) y de Tabachnick y Fiddell (2000) presenta algunos de los métodos más usados en función de los objetivos de la investigación, de la escala en la que son registradas las variables observadas y de su condición, esto es, su jerarquía dentro de las hipótesis de trabajo y de si están bajo control del investigador o no. Estos métodos se estudian en la mayor parte de los cursos de estadística aplicada en programas de grado y posgrado, pero son sólo una parte de una batería muy extensa de pruebas disponibles para tratar distintas situaciones y por ello estas tablas sólo son orientativas. El libro de Sheskin (2011) en su quinta edición, contiene más de 300 pruebas estadísticas para análisis uni y bivariados.

A consecuencia de la existencia cada vez más frecuente de grandes bases de datos generadas por el instrumental de laboratorio disponible (Kahn, 2011) y el continuo desarrollo de nuevas metodologías, vemos una constante renovación de los métodos aplicados (Wang *et al.*, 2011), algunos con nombres sugestivos como redes neuronales o árboles de decisión entre otros (Niederberger, 1996; Sanogo y Yang, 2004). Ciertos métodos basados en el uso intensivo de los datos, llamados de remuestreo (*resampling*, *bootstrapping*, *jackknife*) han ganado popularidad dado el avance en la capacidad de cálculo de las computadoras personales (Yu, 2003).

La mirada Bayesiana

En los últimos años el enfoque bayesiano ha ganado espacio frente al tradicional o frecuentista, sobre todo en áreas de la medicina y de la economía, pero también en ecología (Reckhow, 1990; Dennis, 1996 y trabajos allí citados) y en agronomía (Mila y Carriquiry, 2004; Edwards y Jannick, 2006; Costes *et al.*, 2006). Para explicar la diferencia con el enfoque clásico o frecuentista seguiremos a Gajewski y Simon (2008). Los métodos bayesianos permiten combinar la información previa con los datos observados. La información previa puede obtenerse de estudios realizados previamente o de opinión experta. Partiendo del teorema de Bayes

Información total = Información histórica + Datos

que en la literatura bayesiana vemos mencionar como "distribución posterior", "verosimilitud (*likelihood*)" y "distribución previa (*prior*)", respectivamente. Vemos que el análisis (posterior) es una combinación del conocimiento del investigador sobre el fenómeno bajo estudio antes de realizar la experiencia con los que se aprende de los datos obtenidos al concretarla. En el análisis clásico

Información total = Datos

la información previa no juega ningún rol, y sólo se tienen en cuanta los datos. Conceptualmente, en ausencia de información previa, ambos modos de analizar brindan los mismos resultados. Dicho de otro modo, el enfoque bayesiano nos permite actualizar el conocimiento existente con nueva información. En teoría, es una alternativa simple a la inferencia estadística conocida a partir de la distribución posterior. Pero en la práctica, esta sólo puede obtenerse analíticamente en casos muy sencillos, y la solución sólo puede obtenerse

computacionalmente usando métodos de simulación (Monte Carlo principalmente), generando muestras de la distribución posterior y estimando los parámetros de interés a partir de las mismas. Dos aspectos cruciales son la elección de la distribución previa y si se alcanza o no la convergencia; en caso de elección de una distribución previa inadecuada o de no llegar a la convergencia, las inferencias son tanto imprecisas como confusas o directamente erróneas (Christensen 2005).

Ambos enfoques, clásico o frecuentista y bayesiano, tiene sus pros y contras. Usando un punto de vista pragmático (Chatfield, 2002) la elección depende del objetivo: si se dispone de información previa que se quiere incorporar al análisis, lo indicado sería usar un método bayesiano, que tiene la ventaja de presentar resultados en términos de probabilidad más accesibles. El enfoque bayesiano permite responder cuestiones específicas mejor que un estimador puntual clásico. Pero si se está interesado en la estimación de parámetros, aún en casos complejos, no es necesario recurrir a métodos bayesianos (SAS Institute, 2008). Pero una mirada más rigurosa (Dennis, 1996) enfatiza que la adopción de uno u otro enfoque implica una distinta filosofía de la ciencia en cada caso, y la adopción o rechazo de ciertos conceptos centrales como el rol de la aleatorización, valores de *p* e intervalos de confianza. Una introducción interesante al uso de métodos bayesianos y su aplicación en epidemiología se encuentra en Greenland (2006; 2007); véase también el artículo de Cressie *et al.* (2009) y la discusión subsiguiente en la revista *Ecological Applications*, y la sección "Bayesian Analysis: Advantages and Disadvantages" en el manual de SAS (2008).

Programas estadísticos

La revolución estadística se ha visto complementada, posibilitada y potenciada por la disponibilidad de programas de computación que permiten aplicar a cualquier conjunto de datos los métodos reseñados. Los hay gratuitos y disponibles en la Web, para uso en bioestadística como el Epi-Info (http://wwwn.cdc.gov/epiinfo/) o el G-Stat (www.g-stat.es/), algunos específicos para métodos no paramétricos (MSTAT, http://mcardle.oncology.wisc.edu/mstat), otros concebidos para las necesidades de programas de mejoramiento genético vegetal (Crop-Stat, www.irri.org), o para investigaciones en genética como el InfoGen en su versión estudiantil (www.info-gen.com.ar, Balzarini y Di Rienzo, 2011). Otros permiten un gran número de análisis a través de comandos organizados en "scripts" y son onerosos como SAS (www.sas.com) o GenStat y AS-REML (http://www.vsni.co.uk/) o gratuitos como R (www.r-project.org/). El programa WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/) es el más usado para aplicar métodos bayesianos. Escapa a esta reseña la mención de los programas específicos para los estudios a nivel molecular por su número y rápida evolución. La variedad de métodos disponibles en los programas estadísticos y la facilidad de su uso, en muchos casos lleva a un análisis impropio y a una interpretación equivocada de la información contenida en los datos (Appleton, 1995c). Para un uso eficiente es necesario reconocer la estructura subyacente en los mismos (Appleton, 1995a; 1995b), y elegir los programas estadísticos para cubrir las necesidades reales de análisis que enfrentamos a partir de un estudio integral de los mismos (Li *et al.*, 2011).

Algunos problemas actuales

Algunos temas de la investigación en genética básica y aplicada y su correspondiente análisis estadístico han recibido un tratamiento intensivo en los últimos 25 años (Carbonell y Bramardi, 2001; Gianola, 2001). El primero de ellos, en un orden arbitrario, es la interacción genotipo x ambiente, que constituye un desafío constante para el desarrollo de nuevas variedades (Annichiarico, 2002). El análisis de series de experimentos a escala regional, nacional, e incluso mundial, en base a métodos uni y multivariados (Yue *et al.*, 1997; Flores *et al.*, 1998) ha dado lugar a un continuo avance en el mejoramiento genético y a establecer nuevos conceptos como el de mega-ambientes (Gauch y Zobel, 1997). Los estudios a nivel molecular (Speed y Zhou, 2009, y otros trabajos incluidos

en ese número de *Statistical Methods in Medical Research*) han generado inmensas bases de datos cuyo análisis requiere desarrollar nuevos métodos (Terwilliger y Göring, 2009; Kahn, 2011; Ogutu *et al.*, 2012). Un aspecto que no puede pasarse por alto es la dificultad en replicar los resultados de este tipo de estudios (Ioannidis et al 2008). La búsqueda de nuevas fuentes de germoplasma y la caracterización de las colecciones de poblaciones y razas implican un adecuado estudio de la diversidad usando distintas técnicas multivariadas (Hu *et al.*, 2000; Carbonell y Bramardi, 2001; Balzarini *et al.*, 2011). El mejoramiento animal también ha generado grandes masas de datos, cuyo análisis requirió el desarrollo de nuevos conceptos como el de BLUPs (predictores insesgados) y de métodos tales como los modelos mixtos, que luego se han generalizado (Pianola, 2001; Littell, 2011). Puede verse que la dirección actual es la integración de información de distintas fuentes y el tratamiento de grandes bases de datos.

Coda

El paso final de toda investigación es la presentación de los resultados y su interpretación en congresos o en trabajos publicados en revistas científicas y luego de divulgación, si cabe. El uso de tablas, gráficos y figuras (Wainer, 1992) permite una comprensión rápida, aunque en la práctica suele estar limitado al empleo de gráficos de barras o de dispersión sin medidas del error (Cooper *et al.*, 2002) generados por hojas de cálculo, con sus limitaciones (Vidmar, 2007). Cada instancia de publicación tiene sus normas propias, pero hay buenas guías que pueden seguirse, sobre todo para la presentación gráfica (Yandell, 2007). La disponibilidad de herramientas y sobre todo de conceptos y guías sobre comunicación efectiva (Pocock *et al.*, 2008) deberían permitirnos presentar claramente nuestros resultados y conclusiones.

OBJETIVO	DATOS ANALIZADOS			
	CONTINUOS	ORDINALES	BINARIOS	SUPERVIVENCIA
Describir un grupo	Media, Desvío	Media, Rango IQ	Proporción	Curva Kaplan-Meier
Comparar con un valor teórico	t Student	Wilcoxon	Chi cuadrado Prueba binomial	?
Comparar dos grupos independientes	t Student muestras independientes	Mann-Withney	Fisher Chi cuadrado	Log-rank test Mantel-Haenzsel
Comparar + de 2 grupos independientes	ANOVA a una vía	Kruskal-Wallis Jonckheere-Terpstra	Chi cuadrado	Regresión de Cox
Comparar 2 grupos relacionados	t Student muestras apareadas	Wilcoxon	McNemar	Regresión de riesgos proporcionales
Comparar + de 2 grupos relacionados	ANOVA a dos vías	Friedman	Q Cochran	?
Asociación entre 2 variables	Correlación Pearson	Correlación Spearman	Coeficientes contingencia	?
Predecir 1 variable	Reg linear/no linear simple/múltiple	Reg no paramétrica Sen-Adichie	Reg logística	Regresión de Cox
Asociación entre + de 2 variables	Correlación canónica	?	Análisis de frecuencias multivía	?
Establecer una estructura	Componentes principales	Análisis de correspondencia		?
	Modelos de ecuaciones estructurales			?
Asignar observaciones a grupos	Análisis discriminante	?	Análisis logit	?
Establecer una clasificación	Técnicas de agrupamiento			?

Tabla 1. Sinopsis de los métodos estadísticos más usados, adaptada de Motulsky (1995) y de Tabachnick y Fiddell (2000). Ver texto para definiciones y métodos alternativos.

CONCLUSIÓN

¿Significa todo esto que debemos conocer y/o emplear todos los métodos? En modo alguno. Después de todo, si entendemos qué implica hablar de media o mediana, desvío (y error estándar) o rango, e intervalo de confianza, y sabemos cuándo usar chi cuadrado, t de Student, regresión linear y ANOVA, podemos entender la mayor parte de los trabajos publicados (Reed *et al.*, 2003; Al-Benna *et al.*, 2010). Pero es necesario que los investigadores sean conscientes de las alternativas disponibles, y elijan la metodología de análisis con criterio amplio.

Citando un manual de 1992 de la Auditoría estadounidense (*US General Accounting Office*), para que una investigación sea fructífera se requiere (1) la comprensión de una amplia gama de métodos de análisis de datos, (2) la planificación temprana de la forma en que se analizarán los datos y la revisión de los métodos a medida que el trabajo avanza, (3) identificar cuáles métodos responden mejor a las preguntas formuladas en el estudio, dados los datos obtenidos, y (4) una vez que el análisis ha finalizado, reconocer cómo las debilidades presentes en los datos y/o en el análisis afectan las conclusiones que se han extraído (traducción libre del autor). Hoy es posible para cualquier investigador poner a prueba mediante simulación la metodología de análisis empleada y otras alternativas (MacArdle y Anderson 2004). Conviene tener presente que no todos los hallazgos y descubrimientos informados son luego confirmados, sea por una planificación indecuada, una técnica experimental imperfecta o un análisis incorrecto (*cf.* Begley y Ellis 2012).

AGRADECIMIENTOS

Mi interés en este tema proviene de sendos cursos dictados por María Inés Oyarzábal y Jorge Mariotti hace tiempo. Una versión previa fue leída y comentada por Valeria Borelli y Valeria Paccapelo, y Elsa Camadro me estimuló a concluirla. Los errores, omisiones y arbitrariedades son responsabilidad mía.

Varias ideas provenen de una presentación a una mesa redonda organizada por Martín Grondona en el seno del Grupo Argentino de Biometría en 2006.

BIBLIOGRAFÍA

Al-Benna S., Al-Ajam Y., Way B., Steinstraesser L. (2010) Descriptive and inferential statistical methods used in burns research. Burns 36: 343-346.

Altman D. (1982) Statistics in medical journals. Stat. Med. 1: 59-71.

Altman D.G., Gore S.M., Gardner M.J., Pocock S.J. (1983) Statistical guidelines for contributors to medical journals. BMJ 286: 1489-1493.

Annicchiarico P. (2002) Genotype x environment interactions. Challenges and opportunities for plant breeding and cultivar recommendations. FAO Plant Production and Protection Paper 174.

Appleton D.R. (1995a, b, c) Pitfalls in the interpretation of studies: I, II, III. J. Roy. Soc. Med. 88: 2-4, 188-190, 241-243.

Bailar J.C., Mosteller F. (1988) Guidelines for statistical reporting in articles for medical journals. Ann. Int. Med. 108: 266-278.

Balding D.G. (2006) A tutorial on statistical methods for population association studies. Nature Reviews Genetics 7: 781-791.

Balzarini M., Di Rienzo J. (2011) InfoGen versión 2011. FCA, Universidad Nacional de Córdoba, Argentina.

Balzarini M., Teich I., Bruno C., Peña A. (2011) Making genetic biodiversity measurable: A review of statistical multivariate methods to study variability at gene level. Rev. FCA UNCuyo 43: 261-275.

Bateson W. (1908) The methods and scope of Genetics. Cambridge University Press. (http://www.esp.org/foundations/genetics/classical/holdings/b/wb-methods-08.pdf)

Begley C.G., Ellis L.M. (2012) Raise standards for preclinical cancer research. Nature 483: 531-533,

Bolker B.M., Brooks M.E., Clark C.J., Geange S.W., Poulsen J.H., Stevens H.H., White J.S.S. (2008) Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecolgy and Evolution 24: 127-135.

Carbonell E.A., Bramardi S.J. (2001) La Estadística en la Investigación Genética Vegetal. XXIX Coloquio de la Sociedad Argentina de Estadística. Neuquén, UNComahue, 16 pp.

Chatfield C. (2002) Confessions of a pragmatic statistician. Journal of the Royal Statistical Society: Series D (The Statistician) 51: 1-20.

Christensen R. (2005) Testing Fisher, Neyman, Pearson and Bayes. The American Statistician 59:121-126.

Churchill G.A. (2002) Fundamentals of experimental design for cDNA microarrays. Nature Genetics 32: 490-495.

Cockerham C.C. (1980) Random and fixed effects in plant genetics. Theoretical and Applied Genetics 56: 119-131.

Cohen M.E. (2001) Analysis of ordinal dental data: Evaluation of conflicting recommendations. J. Dent. Res. 80 (1): 309-313.

Cooper R.J., Schriger D.L., Close R.J.H. (2002) Graphical literacy: The quality of graphs in a large-circulation journal. Ann. Emer. Med. 40: 317-322.

Costes J.M., Crossa J., Sanches A., Cornelius P.L. (2006) A Bayesian approach for assessing the stability of genotypes. Crop Sci. 46: 2654-2665.

Cox D.R. (1972) Regression models and life tables (with Discussion). J. R. Stat. Soc. A. 148: 82-117.

Cressie N.A., Calder C.A., Clark J.S., Van Hoef J.M., Wikle C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19: 553-570.

Dennis B. (1996) Discussion: Should ecologists become Bayesians?. Ecological Applications 6: 1095-1103.

Dudley J.W., Moll R.H. (1969) Interpretation and use of estimates of heritability and genetic variances in plant breeding. Crop Sci. 9: 257-262.

Dyke G.V. (1997) How to avoid bad statistics. Field Crops Research 51 (3): 165-187.

Edmondson R.N. (2005) Past developments and future opportunities in the design and analysis of crop experiments. Journal of Agricultural Science 143: 27-33.

Edwards J.D., Jannick J.L. (2006) Bayesian modeling of heterogeneous error and genotype x environment interaction variances. Crop Sci. 46: 820-833.

Eskridge K.M. (2009) Field Trial designs in plant breeding. The Illinois maize breeding and genetics laboratory. http://imbgl.cropsci.illinois.edu/school/presentations/2009/Eskridge.pdf.

Fernández G. (1992) Residual analysis and data transformations: Important tools in statistical analysis. HortScience 27: 297-300.

Fisher R.A. (1918) The correlations between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52: 399-433.

Fisher R.A. (1930) The genetical theory of natural selection. Oxford Clarendon Press, Oxford.

Fisher R.A. (1952) Statistical methods in genetics. Heredity 6: 1-12.

Fisher R.A. (1960) The design of experiments, Seventh Edition. Oliver and Boyd, Edinburgh.

Flores F., Moreno M.T., Cubero J.I. (1998) A comparison of univariate and multivariate methods to analyze environments. Field Crops Res. 56: 271-286.

Gajewski B.J., Simon S.D. (2008) A one-hour training seminar on bayesian statistics for nursing graduate students. The American Statistician 62: 190-194.

Garrett K.A., Madden L.V., Hughes G., Pfender W.F. (2004) New applications of statistical tools in plant pathology. Phytopathology 94: 999-1003.

Gauch H.G., Zobel R.W. (1996) Optimal replication in selection experiments. Crop Sci. 36: 838-843.

Gauch H.G., Zobel R.W. (1997) Identifying mega-environments and targeting genotypes. Crop Sci. 37: 311-326.

Geng S., Hills F.J. (1978) A procedure for determining numbers of experimental and sampling units. Agronomy Journal 70: 441-444.

Gianola D. (2001) Statistics in animal breeding. In: Raftery A.E., Tanner M.A., Wells M.T. (Eds) "Statistics in the 21st Century". Chapman&Hall/CRC, Boca Raton, pp. 34-41.

Gianola D. (2002) Los métodos estadísticos en el mejoramiento genético. In: Cardellino R., Cardellino R. (Eds.) "Genética Animal: Contribuciones en homenaje al Profesor Ing. Agr. Jaime Rovira". Hemisferio Sur, Montevideo, pp. 61-90. http://www.ansci.wisc.edu/facstaff/Faculty/pages/gianola/genetic_improvement.pdf

Gould W.R., Steiner R.L. (2003) Viewpoint: Improving range science through the appropriate use of statistics. J. Range Manage. 55: 526-529.

Greenland S. (2006) Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int. J. Epidemiol. 35: 765-775.

Greenland S. (2007) Bayesian perspectives for epidemiological research: II. Regression analysis. Int. J. Epidemiol. 36: 195-202.

Hacking I. (2007) The laboratory style of thinking and doing. A lecture at the Science, Technology and Society Workshop, National Tsing Hua University Monday 12th November 2007 (Accesed 04/05/09).

Hager W. (2000) About some misconceptions and the discontent with statistical tests in psychology. Methods of Psychological Research Online, 5: 1-31. http://www.psychologie.de/fachgruppen/methoden/mpr-online/issue9/art1/hager.pdf

Hanson W.D. (1959) Minimum Family Sizes for the Planning of Genetic Experiments. Agronomy Journal 51: 711-715.

Holland J.B., Nyquist W.E., Cervantes-Martinez C.T. (2003) Estimating and interpreting heritability for plant breeding: An update. In: Janick J. (Ed.) Plant breeding reviews. Wiley, New York, Vol. 22, pp. 9-111.

Hu J., Zhu J., Xu H.M. (2000) Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. Theor. Appl. Genet. 101: 264-268.

Hua X., Spilke J. (2011) Variance–covariance structure and its influence on variety assessment in regional crop trials. Field Crops Research 120: 1-8.

Ioannidis J.P.A., Allison D.B., Ball C.A., Coulibaly I., Cui X, Culhane A.C., Falchi M, Furlanello C., Game L., Jurman G., Mangion J., Mehta T., Nitzberg M, Page G.P, Petretto E., van Noor V.(2012) Repeatability of published microarray gene expression analyses. Nature Genetics 41:149-155.

Kahn S.D. (2011) On the future of genomic data. Science 331: 728-729.

Kearsey M.J., Farquhar A.G.L. (1998) QTL analysis in plants; where are we now?. Heredity 80: 137-142.

Lawson A.B., Cressie N. (2000) Spatial statistical methods for environmental epidemiology. In: Handbook of Statistics, Vol 18, pp. 357-396. Elsevier, Holanda

Li B., Lingsma H.F., Steyerberg E.W., Lesaffre E. (2011) Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. BMC Medical Research Methodology 11: 77 (http://www.biomedcentral.com/1471-2288/11/77, accessed 26/03/2011).

Lin C.S., Binns M.R., Lefkovitch L.P. (1986) Stability analysis: Where do we stand?. Crop Sci. 26: 894-900.

Littelll R.C. (2011) The evolution of linear models in SAS: A personal perspective. Paper 325-2011, SAS Global Forum 2011.

MacArdle B.H., Anderson M.J. (2004) Variance heterogeneity, transformations, and models of species abundance: a cautionary tale. Can. J. Fish- Aquat. Sci. 61:1294-1302.

Machado S., Petrie S.R. (2006) Symposium-Analysis of unreplicated experiments-Introduction. Crop Sci. 46: 2474-2475.

Maindonald J.H. (1984) Use of statistical evidence in some recent issues of DSIR agricultural journals. NZ J. Agric. Res. 27: 597-610.

Mather K. (1938) The measurement of linkage in heredity. Methuen, London.

Matson P., Potvin C., Travis J. (1993) Statistical methods: An upgrade for ecologists. Ecology 74: 161-161.

McLean R.A., Sanders W.L., Stroup W.W. (1991) A unified approach to mixed linear models. The Am. Stat. 45: 54-64.

Mead R. (1990) The design of experiments. Cambridge Univ. Press. Cambridge.

Mila A.L., Carriquiry A.L. (2004) Bayesian analysis in plant pathology. Phytopathology 94: 1027-1030.

Moll R.H., Robinson H.F. (1967) Quantitative genetic investigations of yield of maize. Der Züchter 37: 192-205.

Montana G. (2006) Statistical methods in genetics. Briefings in Bioinformatics 7: 297-308.

Motulsky H.J. (1995) Intuitive biostatistics. Oxford University Press, New York.

Nelson L.A., Rawlings J.O. (1983) Ten common misuses of statistics in agronomic research and reporting. J. Agron. Educ. 12: 100-105.

Niederberger C. (1996) Computational tools for the modern andrologist. J. Androl. 17: 462-466.

Nelsen T.C. (2002) The state of statistics in agricultural science. Journal of Agricultural, Biological, and Environmental Statistics, 7: 313-319.

Ogutu J.O., Schulz-Streeck T., Piepho H.P. (2012) Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. BMC Proceedings 2012, 6 (Suppl 2): S10.

Onofri A., Carbonell E.A., Piepho H.P., Mortimer A.M., Cousens R.D. (2010) Current statistical issues in Weed Research. Weed Research 50: 5-24.

Paterson S, Lello J (2003) Mixed models: getting the best use of parasitological data. Trends in Parasitology 19:370–375.

Payne R.W. (2006) New and traditional methods for the analysis of unreplicated experiments. Crop Sci. 46: 2476-2481.

Piepho H.P., Büchse A., Emrich K. (2003) A hitchhiker's guide to the mixed model analysis of randomized experiments. Journal of Agronomy and Crop Science 189: 310-322.

Piepho H.P., Büchse A., Richter C. (2004) A mixed modelling approach for randomized experiments with repeated measures. Journal of Agronomy and Crop Science 190: 230-240.

Platt R.W. (1997) Logistic regression and odds ratios. Inj. Prev. 3: 294.

Platt R.W. (1998a) ANOVA, t tests, and linear regression. Inj. Prev. 4: 52-53.

Platt R.W. (1998b) Exploratory analysis: what to do first. Inj. Prev. 4: 140.

Pocock S., Travison T., Wruck L. (2008) How to interpret figures in reports of clinical trials. BMJ 336 (7654): 1166-1169.

Potvin C., Roff D. (1993) Distribution-free and robust statistical methods: Viable alternatives to parametric statistic. Ecology 74: 1617-1628.

Potvin C., Travis J. (1993) Concluding remarks: A drop in the ocean… Ecology 74: 1674-1676.

Quackenbush J. (2002) Microarray data normalization and transformation. Nature Genetics 32: 496-501.

Reckhow K.H. (1990) Bayesian inference in non-replicated ecological studies. Ecology 71: 2953-2059.

Reed J.F., Salen P., Bagher P. (2003) Methodological and statistical techniques: What do residents really need to know about statistics? J. Med. Syst. 27: 233-238.

Rosa G.J., Steibel J.P., Tempelman R.J. (2005) Reassessing design and analysis of two-colour microarray experiments using mixed effects models. Comp. Funct. Genomics 6 (3): 123-31.

Sanogo S., Yang X.B. (2004) Overview of selected multivariate statistical methods and their use in phytopathological research. Phytopathology 94: 1004-1006.

SAS Institute Inc (2008) Introduction to Bayesian Analysis Procedures. In: SAS/STAT 9.2 User´s Guide, Ch. 7. SAS Institute, Inc., Cary, NC. pp. 141-179.

Scheiner S.M. (1993) Introduction: Theories, Hypotheses, and Statistics. In: Scheiner S.M., Gurevitch J. (Eds.) Design and Analysis of Ecological Experiments, Chapman and Hall, New York, NY, pp. 1-13.

Scheiner S.M. (2010) Toward a conceptual framework for biology. The Quarterly Review of Biology 85: 293-318.

Sheskin D. (2011) Handbook of parametric and nonparametric statistical procedures Fifth ed. Chapman & Hall/CRC, Boca Raton, Florida.

Speed T., Zhao H. (2009) Microarrays. Stat Methods Med. Res. 18 (6): 531-532.

Stevens S.S. (1946) On the theory of scales of measurement. Science, New Series, Vol. 103: 677-680.

St-Pierre N.R. (2007) Design and analysis of pen studies in the animal sciences. J. Dairy Sci. (E. Suppl) 90: E87-E99.

Tabachnick B.G., Fiddell L.S. (2000) Using Multivariate Statistics, Fourth Edition. Allen and Bacon, New York.

Tempelman R.J. (2009) Assessing experimental designs for research conducted on commercial dairies. Journal of Dairy Science 92: 1-15.

Terwilliger J.D., Göring H.H.H. (2009) Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. Human Biology 81: 663-728.

US General Accounting Office (1992) Quantitative data analysis: An introduction. Transfer paper 10.1.11. (http://www.gao.gov/assets/80/76118.pdf).

van Putten B., Knippers T., Buurman P. (2010) On design and statistical analysis in soil treatment experiments. Soil Science 175: 519-529.

Velleman P.F., Wilkinson L. (1993) Nominal, ordinal, interval, and ratio typologies are misleading. Am. Statistician 47: 65-72.

Vidmar G. (2007) Statistically sound distribution plots in Excel. Metodološki Zvezki 4: 83-98.

Wainer H. (1992) Understanding graphs and tables. Educational Researcher 21: 14-23.

Wang D., Eskridge K.M., Crossa J. (2011) Identifying QTLs and epistasis in structured plant populations using adaptive mixed lasso. Journal of Agricultural, Biological, and Environmental Statistics 16: 170-184.

Warren W.G. (1986) On the presentation of statistical analysis: reason or ritual. Can. J. For. Res. 16: 1185-1193.

White T.L. (1984) Designing nursery experiments. In: Duryea M.L., Landis T.D. (Eds.) Forest nursery manual: Production of bareroot seedlings. Martinus Nijhoff/Dr W. Junk Publishers. The Hague/Boston/Lancaster, for Forest Research Laboratory, Oregon State University. Corvallis. 386 pp. Chapter 28.

Wilkinson L., APA Task Force on Statistical Inference (1999) Statistical methods in psychology journals: guidelines and explanations. American Psychologist 54: 594-604.

Yandell B.S. (2007) Graphical data presentation, with emphasis on genetic data. Hort. Science 42: 1047-1051.

Yu C.H. (2003) Resampling methods: concepts, applications, and justification. Practical Assessment, Research & Evaluation 8 (19). Accessed December 1, 2005.

Yue G.L., Roozeboom K.L., Schapaugh W.T., Liang G.H. (1997) Evaluation of soybean cultivars using parametric and nonparametric stability estimates. Plant Breed. 116: 271-275.