

algunos de ellos también sean necesarios al efectuar estimaciones). Como consecuencia de esta controversia, lo importante es que debe tenerse muy claro el objetivo de la investigación reconociendo que si lo único que se pretende es conocer un parámetro de la población a través del estudio de la muestra, debe obtenerse un intervalo de confianza de ese parámetro. Por otro lado, si el interés es poner en cuestión una hipótesis de trabajo, debe utilizarse la teoría de pruebas de hipótesis. Lo que suele ocurrir muchas veces es que las hipótesis son tan obvias o de tan poco interés, que no merecen ser verificadas y simplemente debe procederse a estimar la magnitud de las diferencias entre los tratamientos.

En el contexto del excesivo énfasis que se ha dado a las pruebas de significación, indudablemente los problemas más frecuentes ocurren en el uso de la pruebas de comparación múltiple, a pesar de que su uso inadecuado ha sido señalado por varios autores desde hace al menos dos décadas (O'Neill y Wetherill, 1971, Chew, 1976; Petersen, 1977; Little, 1978, 1981; Morse y Thompson, 1981; Carmer y Walker, 1982; Madden, Knoke y Louie, 1982; O'Brien, 1983; Dawkins, 1983; Bryan-Jones y Finney, 1983; Saville, 1990; Lowry, 1992). Esta es un área abierta a la controversia, véase por ejemplo la interesante discusión sostenida por Jones (1984), Perry (1986), Jones y Matloff (1986) y Schultz (1989). Las posiciones al respecto cubren un amplio rango, desde quienes consideran que tienen "un papel único y propio en el análisis de datos experimentales" (Schultz, 1989) hasta quienes afirman que "no son útiles aún cuando sean técnicamente válidos" (Perry, 1986) o de manera más radical "los métodos de comparación múltiple no tienen lugar en la interpretación de los datos" (Nelder, 1971). Petersen (1977), analizó el uso de pruebas de comparación múltiple en el volumen 67 de la prestigiosa revista "*Agronomy Journal*". En el 70% de los casos el uso de las pruebas fue totalmente inapropiado o había métodos alternativos más adecuados. Por su parte, Johnson y Berger (1982), encontraron que en la no menos conocida "*Phytopathology*", volúmenes 70 y 71, el 66% de casi 200 cuadros y figuras abusaron

en la aplicación de aquellas pruebas. Más recientemente y en una especialidad diferente, Lowry (1992) estudió el uso de las pruebas de comparación múltiple en los artículos publicados en las secciones de nutrición de rumiantes y no rumiantes del "*Journal of Animal Science*", de los años 1981 y 1987, encontrando que sólo en el 41 y 52% de los trabajos de esos años, respectivamente, se usó adecuadamente esa metodología estadística.

De esas pruebas, la más utilizada es la propuesta por Duncan (1955). En las revistas "*Environmental Entomology*" y "*Journal of Economic Entomology*" correspondientes a seis meses de 1981, por ejemplo, esa prueba se usó, respectivamente, en el 69% y 83% de los casos, en tanto pruebas alternativas no alcanzaron más del 6% (Jones, 1984). Igualmente importante fue su uso en "*Phytopathology*", volumen 70, en el cual el 65% de los casos en que se usó alguna prueba de comparación múltiple, la prueba de Duncan fue la escogida (Madden, Knoke y Louie, 1982). Haciendo referencia de nuevo al volumen 15 de "*Agronomía Costarricense*", en el 30% de los casos se utilizó el análisis de varianza seguido de una prueba de comparación múltiple, lo que es una proporción elevada si se tiene presente que en el 43% se omitió o fue confusa la descripción del análisis y que el 9% de los trabajos, por su carácter demostrativo, descriptivo, microfotográfico, etc., no requería, aparentemente, un análisis estadístico riguroso.

Los abusos más importantes de las pruebas de comparación múltiple en la investigación agrícola son la comparación de medias de tratamientos en experimentos factoriales, la comparación de la respuesta media a diferentes niveles de un factor cuantitativo y el análisis de tratamientos cuantitativos cuando se tiene interés, previo al experimento, en combinaciones lineales de medias de tratamientos (contrastes). Lo procedente en esos casos es la partición del cuadrado medio de tratamientos o de los grados de libertad de tratamientos en comparaciones de interés, lo que puede hacerse antes del experimento y no sugerido por los datos, o el uso de técnicas de regresión,

según corresponda a las características de los datos (Chew, 1976; Petersen, 1977; Morse y Thompson, 1981; Johnson y Berger, 1982; Dawkins, 1983; Bryan-Jones y Finney, 1983; Perry, 1986; Lowry, 1992). Al respecto es sorprendente la experiencia de Little (1981) en cursos de actualización y seminarios para investigadores agrarios; casi todos los participantes sabían como calcular el estadístico de la prueba de Duncan, pero menos del 10% sabía como descomponer la suma de cuadrados de tratamientos en comparaciones útiles.

Las pruebas de comparación múltiple son apropiadas, según los autores que aceptan su uso, para el análisis de experimentos exploratorios con tratamientos no estructurados, consistentes de materiales no relacionados, variedades o productos químicos por ejemplo, en los que el investigador tiene exactamente la misma actitud hacia y el mismo interés en cada par de tratamientos (Petersen, 1977; Little, 1978; Bryan-Jones y Finney, 1983; Ball, 1992).

Otra consecuencia del énfasis en las pruebas de significación es la costumbre de condicionar las comparaciones específicas entre medias de tratamientos a que el cuadrado medio de éstos sea significativo. Hay que tener en cuenta que ésta es una prueba global, más débil que una prueba para una comparación determinada (Morse y Thompson, 1981; O'Brien, 1983); además se corre el riesgo de no declarar significativa una diferencia real debido a que en el análisis global la significación es "diluida" por las demás diferencias, no significativas del experimento (Chew, 1976; Little, 1981). Diferentes autores recomiendan o desaconsejan esta práctica. Si se trata de una comparación definida a priori, podría ser válido no condicionarla a la significación previa de la prueba F. Sin embargo, si es una comparación sugerida a posteriori por los datos, esta práctica es peligrosa ya que podría declararse significativa una diferencia que realmente no fuera. El caso más extremo es la comparación entre la media más alta y la más baja. En este caso, es de esperar que siendo realmente iguales, por simple azar, fueran declaradas diferentes con una probabilidad mayor que la dada por el nivel

de significación.

Un aspecto importante que debe considerarse en la interpretación de los resultados cuando se usan pruebas de comparación múltiple y que no siempre se tiene en cuenta, es que la magnitud de los valores críticos, por encima del cual se declaran significativas las diferencias, varían para cada una de las pruebas y en consecuencia los resultados y las conclusiones también lo hacen (Madden, Knoke y Louie, 1982). Esta diferencia en el valor crítico es lo que ha dado lugar a las discrepancias entre los estadísticos a la hora de recomendar una u otra prueba. Por tanto, de acuerdo con los objetivos de la investigación, debe saberse cuando no debe emplearse una prueba dada. Una interesante crítica a la mayoría de las pruebas de comparación múltiple la hace Saville (1990), que introduce el concepto de "consistencia". Una prueba es consistente si la decisión depende solamente de la diferencia entre las dos medias, del error típico de esa diferencia, de los grados de libertad del error y del nivel de significación escogido. De acuerdo con eso, las pruebas más frecuentemente utilizadas son "inconsistentes" ya que la decisión también depende del número de poblaciones involucradas en el experimento y de las medias de esas otras poblaciones y en consecuencia el autor citado recomienda el uso de la diferencia mínima significativa no protegida, que es consistente, cuando las características del experimento aconsejan el uso de una prueba de comparación múltiple.

Un error común en el análisis de experimentos factoriales, es que se consideran primero los efectos principales y luego las interacciones, cuando debe ser a la inversa. Dependiendo del tipo e importancia de la interacción (matemática o práctica), deben evaluarse los distintos niveles de un factor dentro de cada nivel del otro factor y no como sucede a veces, que se evalúa la respuesta media de los niveles de un factor, calculada considerando globalmente todos los niveles del otro factor (Johnson y Berger, 1982; Lauckner, 1989). Por ello, debe ponerse mucho cuidado y estudiar a fondo la trascendencia práctica de las interacciones antes de considerar el estudio de

los efectos principales. Igualmente, en los diseños factoriales desequilibrados debe tenerse muy en cuenta la falta de simetría en la descomposición de la suma de cuadrados existiendo diversas formas de plantear el análisis de los datos. Esto debido a que los resultados de las sumas de cuadrados de un efecto dependen del resto de los efectos ya introducidos en el análisis. Este problema nace de la dependencia de los efectos como consecuencia del desequilibrio y es muy similar al que se presenta en el análisis de regresión cuando las variables regresoras están correlacionadas.

El uso inadecuado del análisis de regresión es señalado en relación a estudios de suelos por Webster (1989) y en fitopatología por Johnson y Berger (1982), aunque seguramente el problema es común a otras especialidades de las ciencias naturales. En este tipo de estudios lo que simplemente se obtiene es una ecuación que "mejor" (de acuerdo con criterio determinado de optimización) se ajusta a unos datos de modo que como señalan Ostle y Mensing (1975, pag 166), simplemente porque se haya supuesto una relación funcional particular entre variables (modelo) y se haya seguido un proceso de cálculo determinado (optimización), no se puede pretender que siempre exista esa relación causal entre las variables; es decir, sólo porque se haya encontrado una función matemática que se ajuste bien a los datos, no se está necesariamente en situación de inferir con certeza que un cambio en una variable "causa" un cambio en otra variable. Frecuentemente cuando se analizan variables cuantitativas y su relación, se intenta ajustar modelos mediante técnicas de regresión, a pesar de que las variables en cuestión involucren errores derivados del muestreo y de la precisión de la medida. Solo cuando los predictores se conocen sin error es totalmente apropiado el análisis de regresión (Webster, 1989). En este tipo de análisis la inclusión de los errores típicos de los coeficientes o el coeficiente de determinación ajustado es infrecuente, mientras que el coeficiente de determinación ( $R^2$ ) es el indicador más común de la bondad del ajuste. Sin embargo un  $R^2$  elevado puede enmascarar un modelo inapropiado que puede detectarse estudiando los residuos o graficando las observaciones y los valores

estimados (Carbonell et al 1983). También puede obtenerse un engañosamente alto coeficiente de determinación cuando el modelo está sobreparametrizado (Lauckner, 1984) o cuando se trabaja con los valores medios de la variable en vez de con los valores observados. Otro error muy frecuente en este tipo de estudios es utilizar un modelo de regresión en datos que están estructurados mediante uno o más factores de clasificación (sexo, edad, etc.). En estos casos es necesario comprobar previamente que los respectivos modelos de regresión dentro de cada estructura son iguales entre sí. De la misma forma que para el caso del análisis de varianza, el análisis de regresión requiere una serie de suposiciones básicas que raramente se verifica su cumplimiento. Dentro del contexto de estudiar la relación entre variables, existe una gran confusión entre regresión y correlación. Muchos investigadores utilizan ambos análisis de forma indistinta sin darse cuenta que ambos persiguen objetivos diferentes aunque empleando una metodología similar. Si lo que se pretende es estudiar el grado de asociación entre dos variables resulta inadecuado establecer una ecuación de regresión que sirva para estudiar el comportamiento (o incluso predecir) de la variable dependiente al variar la independiente. En este tipo de estudios sobre la asociación entre variables es muy frecuente caer en la tentación de establecer relaciones de causalidad como consecuencia de los resultados obtenidos. Es bien conocida la relación entre el número de médicos en Africa y el aumento de las enfermedades infecciosas o entre el número de partos y el de cigüeñas sin que nadie se aventure a establecer una relación causa-efecto entre las variables.

Otro problema común en la interpretación de resultados es el uso indiscriminado del coeficiente de variación -CV-, sobre todo cuando se refiere a variables expresadas en porcentajes, conteo, proporciones, etc. (Preece, 1984; Lauckner, 1989). En casos como el porcentaje de germinación ( $p$ ), por ejemplo, se tiene la misma información que con el porcentaje de no germinación ( $q = 1 - p$ ), pero el CV en ambos casos puede ser muy diferente. Esto es así porque la desviación típica es la misma ( $\sqrt{npq}$ ) pero la media no

(np o nq según el caso que se trate). Consecuentemente, en estos casos el CV no es un indicador adecuado de la precisión del experimento.

En el análisis de experimentos en los que se toman medidas repetidas temporal o espacialmente de las mismas unidades experimentales, suele olvidarse que estas medidas pueden estar correlacionadas en mayor o menor medida. Por tanto, no debe introducirse en modelo el efecto del tiempo o el espacio como un factor más despreciando la presencia de la autocorrelación. A menudo se analizan como un diseño en parcelas divididas, incumpliendo frecuentemente las condiciones para que este análisis sea válido (Littell, 1989). Esas condiciones, según los textos clásicos de métodos estadísticos, son homogeneidad de la varianza para todas las medidas e igual correlación entre todos los posibles pares de medidas. Este autor menciona que esas restricciones son muy amplias y lo que se requiere es la denominada condición Huynh-Feldt, a la que se refiere extensamente en su artículo.

Algunas de las técnicas de análisis estadístico cuyo uso es limitado en la investigación agraria, a pesar de sus ventajas con respecto a las alternativas de uso más frecuente, son el análisis de covarianza (Lauckner, 1984), los métodos de análisis multivariado (Pearce y Gilliver, 1978) y el análisis "cluster" (Madden, Knoke y Louie, 1982). En muchos casos esto provoca que, al estudiar las variables una a una y no en conjunto, no se obtenga de los datos experimentales toda la información disponible y que la interpretación de los resultados sea omisa o parcial ya que no considera la correlación entre las variables estudiadas. Con respecto al análisis de covarianza, García de Siles et al (1979) muestran las diferentes conclusiones que se pueden obtener al contabilizar o no la influencia de variables concomitantes en un experimento sobre nutrición animal. Los resultados de los análisis dieron lugar a conclusiones erróneas si se utilizaba un análisis de varianza en vez de uno de covarianza y si no se prestaba atención a las premisas implícitas en este tipo de análisis. Aunque los análisis multivariados tienen un gran interés en la Investigación Agraria, debe resaltarse que

sus suposiciones básicas son aún más restrictivas que las de los correspondientes análisis univariados (principalmente en el análisis de varianza multivariable o el discriminante) o que las decisiones tomadas sobre el proceso de cálculo a seguir dan lugar a resultados diferentes (por ejemplo según se utilice la matriz de covarianzas o de correlaciones en el análisis de componentes principales)

Finalmente, parece conveniente recoger algunas observaciones respecto al uso de programas computacionales de análisis estadístico, dado el gran acceso de los investigadores a los microcomputadores y la disponibilidad de numerosos "paquetes" comerciales para esas tareas. Las características y limitaciones de esos programas deben de considerarse por la influencia que pueden tener en los resultados (Preece, 1984). Una revisión de los principales programas disponibles es la efectuada por Milliken y Remmenga (1989) quienes proponen las características que debería tener el programa ideal. Respecto a las desventajas en el uso de microcomputadores, esos autores señalan, entre otras, su limitación para manejar bases de datos de cierto tamaño, la tendencia a condicionar el análisis de los datos a las disponibilidades del programa utilizado aún cuando no utilice los métodos óptimos desde un punto de vista estadístico y, por último, la incapacidad de algunos programas para analizar datos desequilibrados, con celdas vacías y diseños experimentales complejos. La utilización de estos programas directamente por los investigadores sin un asesoramiento adecuado por parte de los estadísticos puede dar lugar a errores importantes en los resultados de los análisis. Por ejemplo, algunos paquetes no indican claramente que sus análisis son válidos para factores de efectos fijos o qué tipo de descomposición se efectúa en el caso de diseños factoriales desequilibrados o incluso no avisan de que las medias que proporcionan en un análisis de covarianza no son las ajustadas por lo que las comparaciones de medias posteriores son erróneas. La facilidad de uso de estos paquetes hace que el investigador se lance a tumba abierta a emplearlos sin cuestionarse su utilidad o reflexionar siquiera unos minutos sobre el conjunto de datos

de que dispone y la mejor forma de enfocar su análisis.

### Problemas en la presentación de resultados

La presentación o publicación de resultados es solamente una parte del proceso de investigación y su calidad está en gran parte determinada por las etapas anteriores, desde la planificación y el diseño del experimento hasta el análisis y la interpretación de los resultados. Muchos de los problemas en la presentación, desde una perspectiva estadística, son originados por el diseño y el análisis utilizado.

Dentro de las limitaciones de costo y espacio debe suministrarse al lector la información necesaria sobre el problema que se investiga, los objetivos que se persiguen, la metodología usada y la interpretación de los resultados, de manera que tenga los elementos suficientes para repetir la experiencia o para reinterpretar los datos.

Muy a menudo se omite una descripción adecuada del diseño experimental, de los factores y niveles involucrados y de la estructura de los datos. Una mención del programa computacional usado suele sustituir la referencia detallada a los procedimientos de análisis seguidos y del modelo estadístico que describe el diseño empleado (Finney, 1991). La negativa de algunas revistas, muchas veces sin razón, para incluir el análisis de varianza completo en los artículos, no es compensada por los investigadores con una discusión detallada en la sección correspondiente (Nelson, 1989).

Una omisión importante se refiere a indicadores de la varianza o precisión del experimento. Frecuentemente el valor del error típico o de la desviación típica, que aportan mayor información, es sustituido por asteriscos, superíndices, niveles de probabilidad o por las siglas NS, no significativo (Morse y Thompson, 1981; Preece, 1984; Lauckner, 1989) o peor aún, por unas tablas de medias que incluyen una ristra de letras, a, ab, b, bc, bcd, etc. totalmente inadecuadas (ver Aliaga-Morell et al 1987, como ejemplo de la cantidad de letras que se pueden llegar a poner en un cuadro). Es

común encontrar el valor correspondiente a las medias de tratamientos seguidas del símbolo  $\pm$  y otro valor, sin indicación de lo que significa (Lauckner, 1984; Nelson, 1989). Un excesivo número de dígitos decimales, que sugieren una precisión mayor de la que realmente se tiene, es frecuentemente encontrada en artículos e informes de investigación, cuando más de tres dígitos son raramente necesarios (Little, 1981).

## ORIGEN DE LA PROBLEMATICA

Una problemática tan extensa y variada como la expuesta, difícilmente tiene un origen único y simple. La literatura revisada sugiere múltiples explicaciones, en concordancia con los problemas particulares que aborda y la experiencia concreta de los autores, no obstante, a continuación se esbozan las que parecen ser más importantes y de carácter más general.

Refiriéndose al uso inadecuado de la Biometría en el Tercer Mundo, Preece (1984) señala como una de las causas más importantes, la enseñanza de la estadística con una orientación fundamentalmente matemática, que ha puesto el énfasis en la teoría, en perjuicio de las realidades prácticas y concretas. Similar situación se presenta en los libros relacionados con el tema, en los que, por una parte, abunda la jerga y la parafernalia matemática, y por otra, se hace uso de ejemplos triviales, totalmente alejados de las condiciones reales que enfrenta el investigador en el campo. Como consecuencia, los procedimientos estadísticos se siguen ciegamente, sin espíritu crítico, como recetas de cocina. Sin que sirva de consuelo, el mismo autor aclara que la situación es similar en el llamado mundo desarrollado. Parece necesario poner más énfasis en la comprensión de los conceptos fundamentales que en su derivación matemática, así como dar mayor importancia al uso adecuado de los métodos y a la correcta interpretación de los resultados, por encima de su sustento teórico. Debe buscarse un equilibrio entre una perspectiva de usuario y el rigor matemático.

En áreas tropicales, Lauckner (1989), indica que otra de las causas es la existencia

de un reducido número profesionales especializados en Biometría y métodos estadísticos aplicados a la investigación agrícola. Esto dificulta el acceso de los investigadores a la asesoría especializada y la capacidad de los profesionales en Estadística y Biometría para hacer oír sus opiniones. Esta causa es muy patente en Centro America y en España. La relación entre investigadores agrarios y estadísticos aplicados en España es del orden de más de 150 frente a los alrededor de 40 que existen en el INRA francés o frente al valor deseable de unos 15 o 20.

En relación con lo anterior, no es común que se involucre a biometristas y estadísticos en el proceso de investigación desde las etapas iniciales del mismo, más bien se solicita su colaboración cuando aparecen problemas o contradicciones en la interpretación de los resultados y ya es poco lo que puede hacerse. Por otra parte tampoco resulta intelectualmente motivador para aquellos, involucrarse en los proyectos de investigación de agrónomos, biólogos, etcétera, que desde el punto de vista de un estadístico resultan sencillos y no constituyen un reto para el especialista.

Como corolario, se ha producido la tendencia a tomar como modelo para el diseño, análisis y presentación, resultados publicados por otros autores en revistas especializadas de prestigio, aún cuando contengan errores estadísticos o los objetivos y condiciones del trabajo sean distintos. El hecho de que muchos artículos son aceptados para su publicación aún cuando tengan "ejemplos de una aproximación ritual a la Biometría" (Lauckner, 1989), desestimula la crítica respecto a los métodos usados y la búsqueda de asesoría especializada, a la vez que difunde el uso inadecuado de los métodos estadísticos en la investigación.

#### A MANERA DE CONCLUSIÓN

A pesar de que no se dispone de evidencia cuantificada en nuestros países sobre la problemática planteada, no parece aventurado suponer que la situación debe ser al menos similar a la que se ha descrito en esta revisión. Por la importancia del sector

agrario, la necesidad de usar de manera óptima los siempre escasos recursos dedicados a la investigación y por el papel clave que juegan los métodos estadísticos en ese proceso, el estudio y la solución de los problemas existentes deben ser abordados, de manera conjunta y articulada, por todos los sectores e instituciones involucrados.

Una tarea impostergable es la caracterización de nuestra situación particular, teniendo en cuenta todos los esfuerzos de investigación que se realizan, muchos de los cuales se desconocen, atados a la limitada circulación de la llamada "literatura gris". El primer autor de este trabajo ha iniciado el análisis de la revista "*Agronomía Costarricense*" y el segundo autor, de las distintas series de "*Investigación Agraria*"; pero deben considerarse además las tesis de grado y proyectos de graduación que se realizan en nuestras Universidades, los informes de investigación archivados en numerosas Instituciones y otras fuentes disponibles. La identificación y cuantificación de nuestros problemas nos darán la pauta para definir las acciones futuras.

Parece necesario tender puentes entre los especialistas en Estadística y Biometría y los investigadores del área de la Agricultura y los Recursos Naturales, de manera que puedan encontrar un lenguaje común y comprender mejor los mutuos intereses y necesidades. Debemos encontrar la fórmula para aproximarnos a las realidades concretas de nuestro medio, manteniendo la rigurosidad teórica necesaria. La adecuación a todo ello de nuestros programas de formación profesional y la incorporación de un enfoque interdisciplinario deben ser parte de la respuesta a medio plazo. Un programa de formación y capacitación del actual personal investigador inculcándole la forma de pensar en términos estadísticos y el uso adecuado de las herramientas que pone a su disposición la estadística, seguramente produciría un impacto inmediato. Esos son algunos de los retos que debemos asumir juntos, estadísticos e investigadores.

## BIBLIOGRAFIA CITADA.

18

ALIAGA-MORELL, J.R., CULIAÑEZ-MACIÁ, F. A., CLEMENTE-MARIN, G., PRIMO-YUFERA, E. 1987. Differentiation of rice varieties by electrophoresis of embryo protein. *Theoretical and Applied Genetics*, 74: 224-232.

ANDERSON, V. L. 1970. Restriction errors for linear models (an aid to develop models for designed experiments). *Biometrics* 26: 255-268.

BALL, S. T. 1992 Misuses of multiple range tests in crop science research. *Cereal Research Communications*. 3-4: 287-293.

BRYAN-JONES, J. & FINNEY, D.J. 1983. On an error in "Instructions to Authors". *HortScience* 18(3): 279-282.

CAMACHO, J. 1984. Reproducción de bovinos en Costa Rica: Resultados de investigación y Perspectivas. Sesiones de actualización y perspectivas, VI Congreso Agronómico Nacional, San José, 9-13 julio 1984. Volumen 2: 121-140.

CARBONELL, E.A., DENIS, J.B., CALVO R., GONZALEZ F., PRUÑONOSA, V. 1983. Regresión lineal: Un enfoque conceptual y práctico. Monografías de INIA. 188 pgs.

CARMER, S.G. & WALKER, W.M. 1982. Baby's bear dilemma: a statistical tale. *Agronomy Journal* 74: 122-124.

CHEW, V. 1976. Comparing treatment means: a compendium. *HortScience* 11: 348-357.

DAWKINS, H.C. 1983. Multiple comparisons misused: why so frequently in response curve studies?. *Biometrics* 39: 789-790.

DUNCAN, D.B. 1955. Multiple range and multiple F tests. *Biometrics* 11: 1-42.

FINNEY, D. J. 1991. Letter to the Editor. *Biometric Bulletin* 8: 2-3.

GARCIA DE SILES, J. L., CARBONELL E. A., WILSON, L. L., STRICKLIN R. W. 1979. Utilización de variables concomitantes en el diseño de experimentos

- HAMMER, P.A. 1981. Controlling variability. *HortScience* 16(5): 628-630.
- JOHNSON, S.B. & BERGER, R.D. 1982. On the status of statistics in Phytopathology. *Phytopathology* 72(8): 1014-1015.
- JONES, D. 1984. Use, misuse and role of multiple-comparison procedures in ecological and agricultural entomology. *Environmental Entomology* 13(3): 635-649.
- JONES, D. & MATLOFF, N. 1986. Statistical hypothesis testing in biology: a contradiction in terms. *Journal of Economic Entomology* 79: 1156-1160.
- LAUCKNER, F.B. 1989. Survey of the use of statistics in agricultural research journals. *Tropical Agriculture (Trinidad)* 66(1): 2-7.
- LITTLE, T.M. 1978. If Galileo published in *HortScience*. *HortScience* 13(5): 504-506.
- LITTLE, T.M. 1981. Interpretation and presentation of results. *HortScience* 16(5): 637-640.
- LITTELL, R.C. 1989. Statistical analysis of experiments with repeated measurements. *HortScience* 24(1): 37-40.
- LOWRY, S.R. 1992. Use and misuse of multiple comparisons in animal experiments. *Journal of Animal Science* 70: 1971-1977.
- MADDEN, L.V., KNOKE, J.K. & LOUIE, R. 1982. Considerations for the use of multiple comparisons procedures in phytopathological investigations. *Phytopathology* 72(8): 1015-1017.
- MEAD, R. 1984. Confounded experiments are simple, efficient and misunderstood. *Experimental Agriculture* 20: 185-201.
- MILLIKEN, G.A. & REMMENGA, M.D. 1989. Statistical analyses and the personal computer. *HortScience* 24(1): 45-52.
- MORSE, P.M. & THOMPSON, B.K. 1981. Presentation of experimental results. *Canadian Journal of Plant Science* 61: 799-802.

MORTON, J.R. & WYLLIE, D. 1983. El diseño e interpretación de los experimentos sobre alimentación de animales en los trópicos. *Producción Animal Tropical* 8: 306-326.

NELDER, J.A. 1971. Contribution to discussion of paper by O'Neill and Wetherill. *Journal of the Royal Statistical Society B* 36: 244-246.

NELSON, L.A. 1989. A statistical editor's viewpoint of statistical usage in horticultural science publications. *HortScience* 24(1): 53-57.

O'BRIEN, P.C. 1983. The appropriateness of analysis of variance and multiple comparison procedures. *Biometrics* 39: 787-788.

O'NEILL, R. & WETHERILL, G.B. 1971. The present state of multiple comparison methods. *Journal of the Royal Statistical Society B* 36: 218-255.

OSTLE, B. & MENSING, R. W. 1975. *Statistics in Research*. The Iowa University Press. 596 p.

OYEJOLA, B.A. & MEAD, R. 1982. Statistical assessment of different ways of calculating land equivalent ratios (LER). *Experimental Agriculture* 18: 125-138.

PEARCE, S.C. 1980. Randomized blocks and some alternatives: a study in tropical conditions. *Tropical Agriculture (Trinidad)* 57(1): 1-10.

PEARCE, S.C. 1986. Experimental design: the first sixty years. *Tropical Agriculture (Trinidad)* 63: 95-100.

PEARCE, S.C. 1988. Analysis of data from agricultural experiments. *Tropical Agriculture (Trinidad)* 65: 2-5.

PEARCE, S.C. & GILLIVER, B. 1978. The statistical analysis of data from intercropping experiments. *Journal of Agricultural Science (Cambridge)* 91: 625-632.

PERRY, J.N. 1986. Multiple comparison procedures: a dissenting view. *Journal of Economic Entomology* 79: 1149-1155.

PETERSEN, R.G. 1977. Use and misuse of multiple comparison procedures. *Agronomy Journal* 69: 205-208.