

¿FUERON MANIPULADOS LOS DATOS DE MENDEL? (ARGUMENTOS EN CONTRA)

WERE MENDEL'S DATA MANIPULATED? (COUNTERARGUMENTS)

Fidel Márquez-Sánchez¹ y Jaime Sahagún-Castellanos²

¹Centro Regional Universitario de Occidente. Universidad Autónoma Chapingo. Rosario Castellanos. 2332. Colonia Residencial la Cruz. 44950. Guadalajara, Jalisco. (fidelmqz@hotmail.com). ²Departamento de Fitotecnia. Universidad Autónoma Chapingo. 56230. Chapingo, Estado de México. (jsahagun@correo.chapingo.mx)

RESUMEN

Se discuten las opiniones de varios autores sobre la crítica de R. A. Fisher al ajuste de los datos de Mendel a sus hipótesis sobre la herencia en varios caracteres del chícharo (*Pisum sativum*). Datos demasiado-buenos-para-ser-ciertos, se ha argumentado. Tres autores basan su alegato en la forma inadecuada en que Fisher empleó la prueba de ji-cuadrada para sugerir que Mendel falseó sus datos experimentales. Otro autor señala cómo debe interpretarse el uso de la ji-cuadrada, ya que cuando se usa como lo hizo Fisher, para detectar datos falsificados, cualesquier conjunto de datos aceptables, congruentes con la hipótesis, parecerán demasiado buenos cuando se combinan.

Palabras clave: Datos manipulados, Mendel.

INTRODUCCIÓN

Este ensayo no puede tener las divisiones tradicionales de los escritos científicos o tecnológicos; no es un trabajo de investigación. Se trata de presentar a los lectores de Agricultura, Sociedad y Desarrollo, varios artículos que han aparecido acerca de los trabajos de Mendel, hacer algunos comentarios y concluir que los datos de Mendel no fueron manipulados.

EL RAZONAMIENTO DE FISHER

Al análisis estadístico que R. A. Fisher hizo de los datos obtenidos por Mendel, que lo llevó a concluir que la probabilidad de que éstos puedan obtenerse en una nueva serie de experimentos idénticos de acuerdo con la ji-cuadrada de 7 en 100 000, se le ha llamado la paradoja de los datos demasiado-buenos-para-ser-ciertos (Pilgrim, 1984). Es decir, los datos se ajustaban tan bien a la hipótesis nula, que se podía dudar que fueran ciertos. Otros dos genetistas coincidieron con la argumentación de Fisher.

En genética, la ji-cuadrada se usa para probar la hipótesis de que los resultados de los diferentes tipos de individuos que resultan en una progenie se ajustan a una relación (3:1, 9:3:3:1, etc). El valor calculado de la ji-cuadrada (χ_c^2) es función del número observado

ABSTRACT

The opinions of several authors on R. A. Fisher's criticism to the fit of Mendel's data to his hypotheses on the heredity of several character of peas (*Pisum sativum*) are discussed. These data are too-good-to-be-true, it has been argued. Three authors base their line of arguments on the inadequate way in which Fisher used the chi-square test to suggest that Mendel falsified his experimental data. Another author indicates how the chi-square should be interpreted; since if it is used, as Fisher did, in order to detect falsified data, any set of acceptable data, if congruent with the hypothesis, will appear to be too good when combined.

Key words: Manipulated data, Mendel.

INTRODUCTION

This essay cannot have the traditional sections of scientific or technological papers; it is not a research report. The authors just want to present to the readers of Agricultura, Sociedad y Desarrollo, several published articles about Mendel's work, make some comments, and conclude that Mendel's data were not manipulated.

FISHER'S REASONING

The statistical analysis that R. A. Fisher made of the data obtained by Mendel, which led him to the conclusion that the probability that they might be obtained in a new series of identical experiments in accordance with a chi-square of 7 in 100 000, has been called the too-good-to-be-true paradox (Pilgrim, 1984). That is, the data adjusted so closely to the null hypothesis that one could doubt they were true. Two more geneticists joined Fisher's line of arguments.

In genetics, chi-square is used to test the hypothesis that the results of the different types of individuals resulting in a progeny adjust to a certain relation (3:1, 9:3:3:1, etc). The calculated chi-square value (χ_c^2) is a function based on the observed number of individuals of type i (O_i), and the corresponding expected numbers (E_i), under the assumption that the hypothetical relation is true. The test statistic is

de individuos del tipo i (O_i), y de los números esperados (E_i), bajo el supuesto de que la relación hipotética es cierta. La estadística de prueba es

$$\chi_c^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$$

donde k es el número de tipos que pueden aparecer en la progenie. Entre más pequeño sea el valor de χ_c^2 , más evidencia habrá de que la hipótesis es cierta, y la probabilidad (p) de no rechazar indebidamente la hipótesis nula será mayor.

¿ERA EMPÍRICO EL ENFOQUE DE MENDEL?

Antes de analizar los argumentos de los defensores de Mendel, conviene revisar lo que sus partidarios y antagonistas suponían respecto a su actitud ante la herencia. Monaghan y Corcos (1985a) creen que Fisher (1936), Dunn (1965) y Orel (1971), conocedores del trabajo de Mendel, podían pensar que éste era un teórico de la herencia, es decir, que sabía de antemano los resultados experimentales que obtendría, por lo que los datos podrían ajustarse acercándolos más a lo esperado. (A este respecto se piensa que quien ajustó los datos fue un ayudante de Mendel). Por el contrario Olby (1949) dice que Mendel tenía más interés en estudiar el comportamiento de los híbridos, y que éste no tenía que ver con la naturaleza de la herencia. En este contexto Mendel habría sido un empírico, al ver primero qué obtenía de su experimentación y después formular alguna teoría. Sobre esto, Glass (1963) señala que "...era bastante claro que en el caso Mendel, fue la existencia del inalcanzable e insoluble problema hasta ese entonces, lo que lo encaminó a idear los experimentos que diseñó, y no ninguna construcción conceptual (teoría) con la que se hubiera comprometido o hubiera deseado apoyar", es decir, que no hubiera teorizado al respecto. Sin embargo esto no significa que no tuviera idea, al observar sus resultados, de lo qué estaba sucediendo. Como Mendel mismo señaló, él iba encontrando evidencia de algunas regularidades consistentes. Así, en la introducción de su artículo (Mendel, 1866, traducido en Stern y Sherwood, 1966) establece que el propósito de sus experimentos era seguir el desarrollo de la progenie de los híbridos, ya que no se había formulado una ley aplicable a la formación y desarrollo de éstos. Monaghan y Corcos (1985a) añaden que la pregunta clave, bastante probable, era si los híbridos producirían progenies como éstos (to breed true) o no. Siguiendo en su artículo una amplia presentación y discusión sobre la naturaleza anual estacional de los experimentos de Mendel, sobre la notación empleada en la representación de los genes, dichos autores terminan dándole todo el crédito a Mendel, es decir, que éste no tenía una teoría preconcebida sobre la naturaleza de la herencia para poder ajustar sus datos, sino

$$\chi_c^2 = \sum_{i=1}^k (O_i - E_i)^2 / E_i$$

where k is the number of types that may appear in the progeny. The lower the value of χ_c^2 , the greater the evidence of the hypothesis being true, and the higher the probability (p) of the hypothesis not being improperly rejected.

WAS EMPIRICIST THE MENDEL'S APPROACH?

Before analysing Mendel's defenders arguments, it is convenient to start from what the anti and pro Mendelians assumed with respect to his attitude towards heredity. Monaghan and Corcos (1985a) believe that Fisher (1936), Dunn (1965), and Orel (1971), experts on Mendel's work, might have thought that he was a theorist of heredity, in other words, that Mendel knew beforehand what results he was going to obtain in his experiments, and then the data obtained could be adjusted simply approaching them to what was expected. (With regard to this, it is thought that the data were adjusted by an assistant of Mendel). On the contrary, Olby (1949) says that Mendel more than anything else was interested in studying the behavior of hybrids, and that this had nothing to do with the nature of heredity. In this context, Mendel would have been an empiricist, looking first at what he would obtain from his experiments and formulate later a theory. With respect to this, Glass (1963) point out that "...it was quite clear that in the case of Mendel, it was the existence of the unachievable and up to this moment insoluble problem, which drove him to devise the experiment he designed, and not any conceptual construction (theory), to which he would have committed himself or which he would have wished to support", in other words, he would not have theorized on this matter. However, that does not mean, that when observing his results, he had no idea of what was going on. As Mendel himself indicated, he was finding evidence that some consistent regularities were occurring. Thus, in the introduction of his paper (Mendel, 1866, translated in Stern and Sherwood, 1966), the purpose of his experiments was to go on observing the development of the hybrid progeny, since there had not been formulated a law applicable to their formation and development. Monaghan and Corcos (1985a) add that the key question quite probably was, if hybrids would breed true or not. Continuing in their article a wide presentation and discussion about the seasonal annual nature of Mendel's experiments, on his notation employed in the representation of genes, the abovementioned authors finish giving all the credit to Mendel, that is, that he had not had a preconceived theory on the nature of heredity to be able to adjust his data,

que simplemente era un investigador empírico sobre los híbridos y su descendencia.

LOS DEFENSORES DE MENDEL

En un trabajo posterior, Monaghan y Corcos (1985b) discuten la naturaleza estadística de los resultados de Mendel, que fueron luego analizados por Fisher. Si los datos de Mendel estuvieran manipulados los valores de ji-cuadrada mostrados por Mendel tenderían a ser bajos; ellos muestran datos de siete estudios (dos en la generación F_2 y cinco en la generación F_3 . En ésta, una combinación de los dos caracteres en F_2 y cinco caracteres individuales más) y, aunque para el total la ji-cuadrada para la relación fenotípica 3:1 tiene un alto valor con probabilidades (todos los valores probabilísticos que se asientan en lo sucesivo se refieren al nivel observado de significancia de la estadística de prueba, suponiendo que la hipótesis nula es cierta) entre 0.95 y 0.99 (lo que asombró a Fisher), examinando los resultados en cada carácter no encuentran la tendencia a que fueran más bajos. El valor más bajo de ji-cuadrada fue 0.015 ($0.90 < p < 0.95$), y el más alto 0.606 ($0.30 < p < 0.50$). Al final de los dos experimentos en F_2 parece que Mendel ya tenía un conocimiento significativo de que se obtenía la relación 3:1, por lo que los autores de ese artículo señalan que Mendel quería saber si las plantas permanecían constantes en su progenie (to breed true), lo que reafirmaría la idea del estudio de los híbridos que Mendel deseaba hacer. Por ello autofecundó las plantas y estudió la relación genotípica de los progenitores que exhibían el carácter dominante la cual, como sabemos ahora, sería 2:1, es decir, dos tercios de las plantas segregarían 3:1 y el resto no. Los datos de este experimento muestran ji-cuadradas que varían al azar, con el valor más bajo de 0.00 ($p > 0.99$) y el más alto de 2.01 ($0.10 < p < 0.20$); por esta razón, de nuevo, sus resultados fenotípicos producen una probabilidad alta, entre 0.90 y 0.95.

En los experimentos con dos caracteres, (forma y color del cotiledón), Mendel nunca mencionó la ahora famosa relación 9:3:3:1 (Monaghan y Corcos, 1985a), ya que lo único que quería saber era la estructura genética de las plantas F_2 de sus híbridos. Lo mismo sucede con el estudio para tres caracteres, donde obtiene una probabilidad entre 0.80 y 0.90. En el siguiente estudio Mendel retrocruzó el híbrido doble con el probador recesivo directa y recíprocamente, obteniendo la relación 1/4:1/4:1/4:1/4. La probabilidad de la ji-cuadrada del primer experimento ($AaBb \times aabb$) fue alta ($0.80 < p < 0.90$), y la de su cruce recíproca ($aabb \times AaBb$) también ($0.90 < p < 0.95$), por lo que tal vez, consciente o inconscientemente, Mendel cometió algún sesgo, pero continuó con sus trabajos para confirmar su hipótesis. Por esta razón sembró la siguiente generación de sus

that he simply was an empiric researcher on hybrids and their descendants.

MENDEL'S DEFENDERS

In a later work, Monaghan and Corcos (1985b) discuss the statistical nature of Mendel's results, which were analyzed by Fisher. If Mendel's data were manipulated, those authors say, the values of chi-square presented by Mendel would tend to be low; the authors show their data of seven studies (two in generation F_2 and five in generation F_3 . In the latter, a combination of the two characteristics in F_2 and another five individual characteristics), and though for the total of chi-square for the phenotypic relation 3:1 it has a high value, with probabilities (all probabilistic values stated below refer to the observed significance level of the test statistics, assuming that the null hypothesis is true) between 0.95 and 0.99 (which astonished Fisher), examining the results in each characteristic, the tendency to be lower is not found. The lowest value of chi-square was 0.015 ($0.90 < p < 0.95$), and the highest 0.606 ($0.30 < p < 0.50$). At the end of the two experiments in F_2 , Mendel already seemed to have significant knowledge that the relation 3:1 was obtained, that is why the authors of that article indicate that Mendel at that moment wanted to know whether the plants would remain constant in their progeny (or breed true), which would reaffirm the idea of the study of hybrids, that Mendel wanted to conduct. This way, he self-fertilized the plants and studied the genotypic relation of the parents exhibiting the dominant characteristic, which - as we know now - would be 2:1, that is, two thirds of the plants would segregate, 3:1, and the rest would not. The data of this experiment show chi-squares varying at random, with the lowest value of 0.00 ($p > 0.99$) and the highest of 2.01 ($0.10 < p < 0.20$); therefore, again their phenotypic results produce a high probability, between 0.90 and 0.95.

In the experiments with two characteristics (shape and color of the cotyledon), Mendel never mentioned the now famous relation 9:3:3:1 (Monaghan and Corcos, 1985a), since the only thing he wanted to know was the genetic structure of the F_2 of the hybrids. Something similar happens with the study for three characteristics, where a probability between 0.80 and 0.90 is obtained. In the following study Mendel backcrossed the double hybrid with the recessive tester, directly and reciprocally, obtaining the relation 1/4:1/4:1/4:1/4. The probability of the chi-square of the first experiment ($AaBb \times aabb$) was high, ($0.80 < p < 0.90$), and that of its reciprocal cross ($aabb \times AaBb$) as well ($0.90 < p < 0.95$), may be that is why -deliberately or not- Mendel in a way biased the results, but continued with his work to confirm his hypothesis. Therefore, Mendel sowed the next

cruzas del dihíbrido con el probador doble dominante (cruza directa y recíproca) obteniendo $0.90 < p < 0.95$ y $0.80 < p < 0.85$. En experimentos posteriores añadió dos caracteres más, el color de la flor y la altura del tallo. Con todos los estudios: para uno, dos y tres caracteres, la probabilidad conjunta fue de 0.95 a 0.99, lo que nuevamente es una probabilidad muy alta, como en el caso de la relación fenotípica para un carácter. Con objeto de probar que pudo haber un sesgo inconsciente para el estudio de dos caracteres (forma y color del cotiledón), los autores compararon los resultados de Mendel con los de otros siete investigadores; encontraron que el porcentaje de recesivos era muy parecido al hallado por Mendel, lo que indica que el problema de mala clasificación no era sólo de éste, y por otra parte, que en el caso de dos investigadores sus resultados probabilísticos diferían bastante de los hallados por los otros.

Para reforzar la defensa de los resultados de Mendel, Corcos y Monaghan (1985) introdujeron otros dos argumentos: el de la reducción del número de plantas usadas en la F_3 y el cálculo de Fisher de la homocigosis lograda en esta generación. En la generación F_2 Mendel usó alrededor de 500 plantas; sin embargo, en la F_3 sólo usó 100, sembrando 10 semillas por planta. Ésto fue porque Mendel ya presumía que la relación en F_2 era 3:1, por lo que no consideró necesario tener más plantas. Fisher señalaba que tal reducción causaba que Mendel no obtuviera la relación de heterocigotes a recesivos 2:1, sino 1.8874:1.1126 (relación que se obtiene con base en la probabilidad de que un carácter sea dominante, calculada con la distribución binomial). Por tanto, Corcos y Monaghan calcularon los respectivos números de homocigotes de los cinco experimentos mencionados, según las relaciones propuestas por Mendel y Fisher. De acuerdo con los resultados de este último, los datos de Mendel, según la relación 2:1 se ajustan más a lo esperado, pero no sus resultados más sofisticados, basados en la relación 1.8874:1.1126. De acuerdo con la hipótesis de Fisher los valores de ji-cuadrada variaron de 0.03 a 3.47, para Mendel de 0.00 a 2.01, y para Fisher variaron al azar tanto como se esperaba de la hipótesis de Mendel y, en algunos casos, fueron tan bajos como sus resultados.

En esta serie de estudios no es extraño que Monaghan y Corcos (1985a) concluyan su artículo con la siguiente aseveración de Galileo: "Pero he logrado probar éste y otro hechos, no pocos, ni menos relevantes: y lo que considero más importante, que se ha abierto a esta vasta y más excelente ciencia, de la cual mi trabajo es sólo el inicio, modos y medios con los cuales otras mentes más agudas que la mía, explorarán sus confines más remotos".

Pilgrim (1986), experto en el manejo e interpretación de datos experimentales, considera que la conclusión de Fisher (que los datos de Mendel eran falsos) es errónea.

generation of his dihybrid crosses with dominant double tester (direct and reciprocal cross), obtaining the probabilities 0.90 to 0.95 and 0.80 to 0.85. In subsequent experiments, Mendel added another two characteristics (color of the flower and stem height). With all the studies: for one, two and three characteristics, the joint probability was from 0.95 to 0.99, which again was a very high probability, as in the case of the phenotypic relation for one characteristic. With the purpose of proving that there could have been an unconscious bias for the study of the two characteristics (cotyledon shape and color), the authors compared Mendel's results with those of another seven researchers; they found that the percentage of recessives was very similar to the one found by Mendel, which indicates that the deficient classification was not only his problem; and on the other hand, that in the case of two researchers, the probabilistic results differed considerably from those obtained by the others.

To reinforce the defense of Mendel's results, Corcos and Monaghan (1985) introduced two additional arguments: the reduction of the number of plants used in F_3 , and Fisher's calculation of homozygosis achieved in this generation. In the F_2 generation, Mendel used about 500 plants; in the F_3 generation, however, he used only 100, sowing 10 seeds per plant. This, because Mendel already expected that the relation in F_2 was 3:1, therefore he did not consider necessary to have more plants. As Fisher pointed out, such reduction caused that Mendel would not obtain the relation of heterozygotes to recessives 2:1, but 1.8874:1.1126 (relation obtained on the basis of the probability that one characteristic would be dominant, calculated with the binomial distribution). Therefore, Corcos and Monaghan calculated the respective numbers of homozigotes for the five experiments mentioned above, with the relations proposed by Mendel and by Fisher. According to the results of the latter, Mendel's data depending on the relation 2:1, are better fitted to what is expected, but his more sophisticated results, based on the relation 1.8874:1.1126, are not. In accordance with Fisher's hypothesis, the chi-square values varied from 0.03 to 3.47, for Mendel, from 0.00 to 2.01, and for Fisher they varied randomly as much as expected from Mendel's hypothesis, and in some cases, they were as low as his results.

So it is not surprising that in this study series Monaghan and Corcos (1985a) conclude their article quoting Galileo's assertion: "But this and other facts not few in number or less worth knowing I have succeeded in proving: and what I consider more important, there have been opened up to this vast and most excellent science, of which my work is only the beginning, ways and means by which other minds more acute than mine will explore its remote corners".

Este autor cuestiona los métodos utilizados por Fisher, que luego fueron adoptados por otros autores. Monaghan y Corcos (1985a) concluyeron que a pesar de que tales métodos estadísticos fueron los apropiados, sus resultados (que coincidían con los de Fisher) no eran lógicos, particularmente los que se derivan de la combinación de información obtenida en varios experimentos independientes, no obstante que en lo individual los datos de cada experimento no mostraban sesgo alguno. Ante esta situación, lo que habría que investigar era si el método utilizado por Fisher, basado en la ji-cuadrada, era adecuado para detectar datos falsificados, y si los datos de Mendel habían sido manipulados. Pilgrim (1986) argumenta que el método que se basa en la ji-cuadrada sólo produce, sobresalientemente bien, la probabilidad de que los datos experimentales sean compatibles con la hipótesis, y que Fisher la usó para otro propósito: sabiendo que la hipótesis era correcta y que 105 datos se ajustaban a ella (puesto que consideraba que Mendel sabía lo que debía esperar), Fisher usó la ji-cuadrada para determinar si los datos de Mendel constituían una muestra aleatoria o estaban manipulados. También investigó la limpieza de los datos de Mendel; los métodos de Pilgrim, sin embargo, fueron muy diferentes: calculó la probabilidad exacta de que los datos obtenidos por Mendel fueran los obtenidos. El cálculo lo hizo con la distribución binomial con parámetros n (número de plantas) y p (probabilidad de la ocurrencia de un genotipo o fenotipo, determinado según la hipótesis particular). Así, para el caso de la proporción de dominantes heterocigóticos en la F_2 de un monohíbrido, los parámetros fueron $n=100$ y $p=\text{proporción de los fenotipos dominantes genéticamente heterocigóticos } (2/3)$. La media de esta distribución es $np=100(2/3)=66.666$ y la varianza es $np(1-p) = 100(2/3)(1/3)=100(2/9)=22.22$, con una desviación estándar de 4.71. De acuerdo con el teorema de De Moivre-Laplace, aproximadamente dos tercios de todos los valores (cada valor es el número de heterocigóticos resultantes en una muestra aleatoria de tamaño 100) deberán estar en el intervalo 66.666 ± 4.71 ; es decir, redondeando, en el intervalo [62, 72]. Los valores que obtuvo Mendel en los cinco experimentos en que usó 100 como tamaño de muestra fueron 64, 71, 60, 67 y 72, números muy congruentes con el razonamiento de Pilgrim. Para que los datos de Mendel hubieran sido demasiado buenos, afirma Pilgrim, los cinco valores deberían haber estado en el intervalo [65, 69] o quizá en el [66, 68]. Con los valores obtenidos, Pilgrim concluyó que no hay razón para sospechar que los resultados no corresponden a otra cosa que no sea una muestra aleatoria.

Para evaluar el resto de los datos de Mendel Pilgrim calculó la probabilidad (Pd) de que cada valor estuviera tan o más cercano a la media $[100(2/3)]$ que en los experimentos de Mendel. Por ejemplo,

Pilgrim (1986), experienced in handling and interpreting experimental data, considered Fisher's conclusion (that Mendel's data were false), to be erroneous. This author questioned the methods utilized by Fisher, which later were adopted by other authors. Monaghan and Corcos (1985a) concluded that in spite of such statistical methods being appropriate, his results (which agreed with Fisher's) were not logic, particularly those deriving from the combination of information obtained from various independent experiments, though individually the data of each experiment did not present any bias. Considering this situation, an analysis had to be made to determine if the method utilized by Fisher, based on chi-square, was adequate to detect falsified data, and whether Mendel's data had been manipulated. Pilgrim (1986) argued that the method based on chi-square only produces -remarkably well- the probability that the experimental data would be compatible with the hypothesis, and that Fisher used it for other purposes: knowing that the hypothesis was correct and that 105 data adjusted to it (since he considered that Mendel already knew what to expect), Fisher used chi-square to determine whether Mendel's data represented a random sample or were biased by the later. Pilgrim also examined the cleanness of Mendel's data; Pilgrim's methods, however, were rather different: he calculated the exact probability that the data obtained by Mendel were those obtained. The calculation was made with the binomial distribution, with parameters n (number of plants) and p (probability of the occurrence of a genotype or phenotype determined according to the hypothesis in consideration). Thus, for the case of the proportion of dominants being heterozygous in the F_2 of a monohybrid, the parameters were $n=100$ and $p=\text{proportion of dominant phenotypes genetically heterozygous } (2/3)$. The mean of this distribution is $np=100(2/3)=66.666$, and the variance is $np(1-p) = 100(2/3)(1/3)=100(2/9)=22.22$, with a standard deviation equal to 4.71. According to the De Moivre-Laplace theorem, approximately two thirds of all values (each value is the number of heterozygous resultants in a random sample of size 100) will be in the interval 66.666 ± 4.71 , that is, rounding, in the interval [62, 72]. The values obtained by Mendel in the five experiments, in which he used 100 as sample size were 64, 71, 60, 67, and 72, very reasonable numbers, based on Pilgrim's reasoning. For Mendel's data having been too good, Pilgrim states, the five values should have been in the interval [65, 69] or perhaps in [66, 68]. Because of the values obtained, Pilgrim concluded that there is no reason to suspect that the results correspond to anything else but a random sample.

In order to evaluate the rest of Mendel's data, Pilgrim calculated the probability (Pd) of each value being as close or more to the mean $[100(2/3)]$ as in Mendel's experiments. For example, for the value 64, since the mean

para el valor 64, como la media es 66.667, $Pd = P(X=64) + P(X=65) + P(X=66) + P(X=67) + P(X=68) + P(X=69) + P(X=70)$. De los tres primeros experimentos que involucraron la relación 3:1, en el primero [semilla redonda (R) vs semilla arrugada (A)] ocurrieron los valores 5 474(R) y 1 850(A) para los que $Pd=0.378$, lo que significa que en 37.8% de los casos los valores de un muestreo aleatorio hubieran estado tan o más cercanos de lo esperado que los de Mendel. Para los dos experimentos restantes $Pd=0.113$ y $Pd=0.481$. Estos resultados, del todo normales en un experimento de este tipo, comenta Pilgrim, difícilmente pueden considerarse extraordinarios. Para dar una idea más clara del significado de estas probabilidades, el autor presenta un ejemplo muy sencillo: para $Pd=0.481$ esta probabilidad es aproximadamente la que corresponde a la ocurrencia de dos hijos del mismo sexo en familias con dos hijos. La media de Pd en 14 experimentos de Mendel fue 0.487, lo que significa que 48.7% de las veces los resultados hubieran estado así de cerca o más a su media, o valor esperado. Con esta información Pilgrim tuvo que concluir que no había razón para creer que los datos de Mendel habían sido obtenidos deshonestamente. Una vez que concluyó que los datos no habían sido falseados, quedaba por determinar qué fue lo que hizo que así pareciera desde la óptica Fisheriana.

Para analizar qué tan apropiada fue la metodología utilizada por Fisher, Pilgrim, considerando que si bien los datos de Mendel no fueron "demasiado buenos para ser ciertos", se preguntó ¿por qué la ji-cuadrada usada por Fisher los mostró como tales? Usando esta distribución, sucede que si los datos son compatibles con la hipótesis las probabilidades altas (p) no son raras. También se preguntó Pilgrim ¿cuándo los datos aparecen como demasiado buenos? Al parecer Fisher nunca reflexionó sobre esto; sin embargo, consideró que a medida que p era mayor los datos se hacían más sospechosos. Por lo anterior, Pilgrim creyó necesario obtener información que diera respuesta a estas preguntas. Para ello analizó con ji-cuadrada cuatro grupos de datos de cinco muestras de tamaño 100 cada uno. Los datos serían probados vs la hipótesis de que una mitad serían de un tipo y la otra mitad de otro tipo. Los grupos de datos fueron: 1) datos honestamente derivados (por muestreo aleatorio); 2) datos con una variación excesiva (descritos por Pilgrim como los de un investigador con mala suerte); 3) datos ostensiblemente cargados a un lado de la curva normal, 4) datos demasiado buenos. En el primer grupo obtuvo un valor de $\chi^2=2.32$, con 5 grados de libertad, para una $p=0.80$; en el segundo $\chi^2=10$ y $p=0.5$. Respecto a datos observados (O_i) ostensiblemente cargados a un lado de la normal (grupo 3), que no se ajustan a la hipótesis, la prueba no discrimina entre datos a un

is 66.667, $Pd = P(X=64) + P(X=65) + P(X=66) + P(X=67) + P(X=68) + P(X=69) + P(X=70)$. Of the three first experiments involving the relation 3:1, in the first [round seed (R) vs wrinkled seed (A)], the values 5474 (R) and 1850 (A) for which $Pd=0.378$, which means that in 37.8 % of the cases the values of a random sampling would have been as close to, or closer to what was expected of what Mendel's data were. For the two remaining experiments, $Pd=0.113$ and $Pd=0.481$. These results, totally normal in an experiment of this type, Pilgrim comments, can hardly be considered extraordinary. To make the significance of these probabilities more obvious, the author gives a very simple example: for $Pd=0.481$, the probability is approximately the one corresponding to the occurrence of two children of the same sex in families of two children. The mean of Pd in 14 of Mendel's experiments was 0.487, which means that in 48.7% of the cases the results would have been as close or closer to its mean or expected value. With this information, Pilgrim had to conclude that there was no reason to believe that Mendel's data had been obtained dishonestly. After concluding that Mendel's data had not been falsified, there was the question of why this appeared feasible from Fisher point of view.

To analyze how appropriate the methodology utilized by Fisher was, Pilgrim wondered, considering that if Mendel's data were not "too-good-to-be-true", why the chi-square used by Fisher made them look like that. Using this distribution, it happens that if the data are compatible with the hypothesis, high probabilities (p) are not unusual. He also questioned at what moment the data aroused the suspicion of being too good. It seems that Fisher never reflected on this matter. He assumed, however, that the higher p, the more suspicious were the data. Because of this situation, Pilgrim considered necessary to obtain information to answer these questions. To that end analyzed four hypothetical data groups of five samples, size 100 each. The data would be tested vs the hypothesis that one half would be of one type, and the other half of a different type. The data groups were: 1) data honestly derived (by random sampling), 2) data with excessive variation (described by Pilgrim as of a researcher with bad luck), 3) data evidently charged to one side of the normal curve, 4) too-good data. In the first group, a value of $\chi^2=2.32$ was obtained, with 5 degrees of freedom, $p=0.80$; in the second $\chi^2=10$ and $0.1=p=0.5$. With respect to data observed (O_i) obviously charged to one side of the normal (group 3), which do not adjust to the hypothesis, the test does not discriminate between data on one side of the mean (or expected value E_i) or on the other; due to the fact that the statistic used is

$\sum_{i=1}^k (O_i - E_i)^2 / E_i$. In this case the results were:

lado de la media (o valor esperado, E_i) o al otro; debido a que la estadística usada es $\sum_{i=1}^k (O_i - E_i)^2 / E_i$. En este caso los resultados fueron: $\chi^2=2.32$ y $p=0.8$, ¡igual al grupo 1 de datos honestos! Finalmente, en el caso de datos demasiado buenos se obtuvo $\chi^2 = 0.32$ con $p>0.99$.

Para investigar qué pasa cuando se analiza conjuntamente los datos de varios experimentos, como lo hizo Fisher con los datos de Mendel, Pilgrim ideó varios escenarios. Con dos conjuntos de datos como los que hubieran sido obtenidos en muestras aleatorias (como los del grupo 1) se obtuvo: $\chi^2=4.62$ y $p=0.9$. Con tres de estos conjuntos obtuvo un valor de $\chi^2=6.96$ y $p=0.96$; es decir, cuando los datos buenos se repiten, es más probable que los datos confirmen la hipótesis. Pilgrim concluye que así es como se debe interpretar el uso de la ji-cuadrada, y que cuando se usa, como Fisher lo hizo, para detectar datos falsificados, cualquier conjunto de datos aceptable, si es congruente con la hipótesis, parecerá demasiado bueno cuando se combine.

Respecto a la crítica de Fisher de que los datos de Mendel están dentro de una desviación estándar, Pilgrim argumenta, con base en la aproximación de la binomial a la normal, que aproximadamente dos terceras partes de los datos deben estar dentro del error estándar. Él, como cualquier conocedor de la distribución normal, encuentra incomprendible la posición de Fisher, y concluye que la ji-cuadrada no es apropiada para detectar datos falsificados, y que no hay razón para que se cuestione la honestidad de Mendel. Pilgrim no fue el único que cuestionó los métodos estadísticos utilizados por Fisher para dictaminar sobre la honestidad de Mendel. Weilling (1986) concluyó que, como la varianza de las segregaciones genéticas es menor que la del tipo binomial supuesto, la ji-cuadrada no es una estadística adecuada para el propósito que persiguió Fisher. También concluyó que, como las varianzas de las diferentes segregaciones no son homogéneas, no se pueden combinar los datos de diferentes experimentos para ser analizados, como lo hizo Fisher. A continuación se presenta un resumen de los análisis de Weilling para obtener sus conclusiones. Con respecto a la magnitud de las varianzas de las segregaciones genéticas, consideró que en el ovario de una flor de la especie que utilizó Mendel para su estudio de híbridos debe haber cuando mucho 12 óvulos, pero no todos producen semilla. Si la hembra es de genotipo aa y el macho es Aa, y si el número de óvulos fertilizados fuera W, esta variable (números de cigotes resultantes) no seguiría una distribución binomial sino una hipergeométrica. La varianza de la Hipergeométrica es menor que la de la Binomial. Por esta razón concluyó que la distribución binomial que Fisher usó para determinar la relación 1.8874:1.1126, en lugar de la supuesta relación 2:1, no fue adecuada, y

$\chi^2=2.32$ and $p=0.8$, equal to group 1 of honest data! Finally, in the case of too-good data, $\chi^2=0.32$ with $p>0.99$ was obtained.

To find out what happens when the data of several experiments are jointly analyzed, as Fisher did with Mendel's data, Pilgrim devised several scenarios. With two groups of data, like those which would have been obtained in random samples (like those of group 1), he obtained $\chi^2=4.62$ and $p=0.9$. With three of these groups, he obtained $\chi^2=6.96$ and $p=0.96$.

In other words, when good data are repeated, it is more probable that the data confirm the hypothesis. Pilgrim concludes that this is the way how the use of chi-square must be interpreted, and when it is used, as Fisher did, to detect falsified data, any acceptable group of data, if they are congruent with the hypothesis, will seem to be too good when combined.

With respect to Fisher's criticism that Mendel's data are within their standard deviation, Pilgrim argues, based on the approximation of the binomial to the normal, that approximately two thirds of the data must fall within the standard error, Pilgrim, as perhaps any expert on normal distribution, finds Fisher's position incomprehensible, and concludes that chi-square is not suitable to detect falsified data, and that there is no reason to question Mendel's honesty. But Pilgrim was not the only one who questioned the statistical methods utilized by Fisher to judge Mendel's honesty. Weilling (1986) concluded that as the variance of genetic segregations is lower than that of the assumed binomial type, chi-square is not an adequate statistic for Fisher's purpose. He also deduced that, as the variances of the different segregations are not homogeneous, the data of different experiments cannot be combined for being analyzed, as Fisher did, with the chi-square distribution. Next, a summary of the analyses made by Weilling to obtain his conclusions is presented. With respect to the magnitude of the variances of genetic segregations, he thought that in the ovary of a flower of the species, used by Mendel for his study of hybrids, there must be at the most 12 ovules, not all of which produce a seed. If the female is of genotype aa and the male is Aa, and if the number of fertilized ovules were W, this variable (numbers of resulting zygotes) would not follow a binomial distribution, but an hypergeometric. The variance of the hypergeometric distribution is smaller than that of the binomial distribution. Therefore, Weilling concluded that the binomial distribution, used by Fisher to determine the relation 1.8874:1.1126, instead of the assumed relation 2:1, was not adequate, and consequently the criticism at Mendel was not justified either. This 2:1 relation was utilized to determine, in the F_2 of a monohybrid the proportion of dominant phenotypes whose offspring, generated by self-fertilization, was homogeneous or heterogeneous.

en consecuencia la crítica a Mendel tampoco lo fue. Esta relación 2:1 fue utilizada para determinar en la F₂ de un monohíbrido la proporción de fenotipos dominantes cuya prole, generada por autofecundación, era homogénea o heterogénea.

También es interesante cómo Weilling (1986) analizó el supuesto de homogeneidad de varianzas requerido para el uso correcto de la ji-cuadrada para analizar datos provenientes de varios experimentos. Él consideró que en los caracteres de forma y color de semilla la segregación resultante se observa en las semillas que darán lugar a las plantas F₂, mientras que en los de planta la segregación se observará en las que generen las semillas. Como en los procesos de germinación de semillas y desarrollo de plantas pueden ocurrir pérdidas, la varianza para estos caracteres debe ser mayor que para color y forma de semilla. Esto da como resultado que, por la heterogeneidad de varianzas, el uso de la ji-cuadrada para el análisis combinado de datos de diferentes experimentos, como el que hizo Fisher con los datos de Mendel, no sea apropiado.

CONCLUSIONES

De acuerdo con la revisión de los datos de Mendel y de los argumentos que Fisher empleó para descalificar sus métodos, los autores que han opinado al respecto, cada uno con su particular aportación, llegan a la conclusión (sostenida por nosotros) de que los datos mendelianos fueron correctamente analizados. Vale la pena recalcar que en este respecto Fisher pudo estar equivocado.

LITERATURA CITADA

Corcos, A., and F. Monaghan. 1985. More about Mendel's experiments: where is the bias? *The Journal of Heredity* 76: 384.
 Dunn, L. C. 1965. Mendel, his work and his place in history. *Proc. Am. Phil. Soc.* 109:189-198.
 Fisher, R. A. 1936. Has Mendel's work been rediscovered? *Am. Science* 1: 115-137.
 Glass, B. T. 1963. The establishment of modern genetical theory as an example of the interaction of different models, techniques, and inferences. *In: Scientific Change*. A. C. Crombie, Ed. Basic Books Inc. Pt. 6, No. 17: 251-541.
 Mendel, G. 1866. Versuch über Pflanzen-Hybriden. *Verhandlungen des Naturforschenden Vereines in Brünn* 4(1865) Abhandlungen. pp: 3-47.

It is also interesting how Weilling (1986) analyzed the assumption of homogeneity of variances, which is required for the correct use of chi-square for the analysis of combined data coming from various experiments. He considered that in the characteristics of seed shape and color, the resulting segregation is observed in the seeds that bring forth F₂ plants, while in those of plant of the segregation will be observed in the plants that generate the seeds. As in the processes of seed germination and plant development may occur losses, the variance for these characteristics must be higher than that for seed color and shape. The result: because of the heterogeneity of variances, the use of chi-square for the combined analysis of data of different experiments, as Fisher did with Mendel's data, is not appropriate.

CONCLUSIONS

According to the revision of Mendel's data and the arguments employed by Fisher to disqualify Mendel's methods, the authors which have contributed to the discussion, each with his particular contribution, have come to the conclusion (sustained by the authors of this essay), that Mendel's data were correctly analyzed. It is worth emphasizing that Fisher, at least with respect to these studies, may have been wrong.

- End of the English version -

Monaghan, F., and A. Corcos. 1985a. Mendel, the empiricist. *Heredity* 76: 49-54.
 Monaghan, F., and A. Corcos. 1985b. Chi-square and Mendel's experiments: where's the bias? *Heredity* 76: 307-309.
 Olby, R. 1949. Mendel no Mendelian. *Hist. Science* 17: 53-72.
 Orel, V. A. 1971. A reconstruction of Mendel's Pissum experiments and an attempt at an explanation of Mendel's way of presentation. *Folia Mendeliana* no. 6. pp: 41-60.
 Pilgrim, I. 1984. The too-good-to-be-true paradox and Gregor Mendel. *The Journal of Heredity* 75: 501-502.
 Pilgrim, I. 1986. A solution to the too-good-to-be-true paradox and Gregor Mendel. *The Journal of Heredity* 77: 218-220.
 Stern, C., and E. R. Sherwood. 1966. *The Origins of Genetics: A Mendel Source Book*. W. H. Freeman.
 Weilling, F. 1986. What about R. A. Fisher's statement of the too-good data of Mendel's pismus paper? *The Journal of Heredity* 77: 281-283.