

LUIS A.
ROMO
SALTOS



ÁREA HISTÓRICA
DEL CENTRO DE INFORMACIÓN INTEGRAL

USO Y ABUSO DE LA ESTADISTICA

USO Y ABUSO DE LA ESTADISTICA

LUIS A. ROMO S.

Consideraciones Generales.

La experimentación científica es una actividad que se realiza operacionalmente con el fin de comprobar las hipótesis que se plantean en el curso de una investigación. A menudo, se ignora que toda medida que se realiza está viciada de error sin que importe cual sea la exactitud y la precisión que se alcancen.

Los errores aleatorios son fortuitos, tienen igual probabilidad de ser de signo positivo o negativo, son ínfimos y se hallan distribuidos al azar.

Los datos experimentales están siempre, de modo inevitable, viciados de los errores aleatorios que también se denominan experimentales o de medida, pero jamás deben estar cargados de errores sistemáticos.

Los errores experimentales están sujetos al tratamiento estadístico que permite separarlos analíticamente de los datos experimentales con el fin primordial de derivar inferencias válidas. Por ejemplo, cuando se afirma que en una investigación de la estatura de dos grupos de personas, los promedios son 172 cm. y 176 cm., interesa establecer si en realidad estas medidas indican que se trata de una diferencia real, propia de las estaturas de los individuos de los dos grupos, o a lo mejor resulta que es aparente porque los dos promedios están viciados de los errores de medida.

En cambio, los errores sistemáticos son de igual signo, o tienen la propiedad de ser periódicos tal como se observa cuando se mide la viscosidad de un líquido en un laboratorio donde se carece de control de la temperatura. Estos errores se originan por algún defecto de funcionamiento del aparato o instrumento de medida, por calibración defectuosa, o por el uso de patrones de medida que no se ajustan a las exigencias impuestas por convención internacional.

Los errores sistemáticos jamás se pueden eliminar mediante un tratamiento estadístico porque solo se depuran eliminando el origen del vicio. Si un aparato no funciona correctamente, hay que repararlo. Los instrumentos y/o patrones de medida deben ser calibrados en relación a los patrones internacionales de medida, etc.

Hace mucho tiempo, Fisher^a y Snedecor^b se ocuparon de estructurar las bases de la correcta aplicación de la Estadística como el medio para separar analíticamente los errores de los datos experimentales. Los progresos que se han logrado en esta importante materia se

- a. R. A. Fisher, *Statistical Methods for Research Workers*, 11th Ed., Hafner Publishing Co., New York (1950).
- b. George Snedecor, *Statistical Methods*, Iowa Sta. Univ. Press, Ames (1957).

patentizan en varias obras^{d,e,f} que se refieren a los métodos del planeamiento de los experimentos científicos.

No cabe que todavía existan lugares donde se sostiene tácitamente que no importa como se realice la investigación, una vez que la Estadística subsana todo! Se debe reconocer que ningún método estadístico puede servir para mejorar lo que está mal hecho. La experimentación debe ser siempre realizada con una diligencia que asegure la obtención de datos que patenticen la realidad sin que importe que concuerdan o no con tal o cual hipótesis preconcebida. Vale advertir al respecto que cuando una hipótesis no concuerda con los hechos, se la debe desechar.

Estos conceptos básicos, a menudo, no son del dominio de quienes realizan la experimentación científica, tal como se puede deducir de las formas de presentación de los datos y de los análisis estadísticos superficiales y a menudo erróneos que se efectúan.

Reducción del Error Experimental.

En el curso de la experimentación, el objetivo que se persigue es el de obtener resultados que permitan derivar las máximas generalizaciones a base de un mínimo número de experiencias. Con este fin, al realizar los experimentos, se debe siempre aplicar los métodos que

conduzcan a reducir los errores a un mínimo. Se logra este objetivo, mediante el uso de las replicaciones una vez que el error es inversamente proporcional a la raíz cuadrada del número de observaciones. Igualmente, para eliminar la tendencia humana a subordinar la realización de las observaciones a tal o cual plan premeditado, con el fin de reducir a un mínimo la heterogeneidad del material de investigación, se deben siempre aplicar al azar los tratamientos. Por ejemplo, si se debe utilizar 30 cobayos en grupos de 5 en un experimento para investigar los efectos de una droga, se los ha de dividir en los seis grupos, asignándolos uno por uno al azar para de este modo reducir a un mínimo el efecto de la heterogeneidad de este material experimental.

En experimentos con más de dos variables independientes, se alcanza estos objetivos mediante el concurso de la teoría del planeamiento de experimentos científicos.

La estrategia que queda puntualizada de nada serviría si al mismo tiempo no se realiza la experimentación mediante la aplicación de los métodos más depurados porque aunque caigamos en redundancia, ningún método estadístico por excelente que sea, puede substituir a una experimentación realizada defectuosamente.

- c. R. A. Fisher, The Design of Experiments, 6th Edit., Hafner Publishing Co., New York (1951).
- d. H. B. Mann, Analysis and Design of Experiments, Dover Publ., New York (1950).
O. L. Davis, Design and Analysis of Industrial Experiments, Oliver & Boyd, London (1964).
- e. Luis A. Romo S., Métodos de Experimentación Científica, Ed. Casals, Barcelona (1964).
- f. S. Vajda, The Mathematics of Experimental Design, Charles Griffin & Co., London (1967).

Abuso de los Métodos Estadísticos.

1.—Una observación general.

El nivel de los cursos de Estadística que se dictan en algunas instituciones de educación superior bordea la mediocridad y hace de esta importante disciplina una herramienta mecánica que se la aplica sin ningún empeño de hacer un verdadero análisis de errores; pues, parece que el fin que se persigue es el de adornar los datos que se han recogido con las desviaciones típicas sin ninguna sujeción a un plan racional de investigación.

Sorprende que en muchas instituciones de educación superior de América Latina se dicten cursos de Estadística que se reducen a recetarios de fórmulas que se usan ignorando de donde proceden, que limitaciones tienen y en que casos se deben aplicar. Cuán común es constatar que se realizan experimentos para determinar el efecto de tal o cual variable para luego de obtenidos los resultados tratar de acomodarlos a tal o cual tratamiento estadístico.

En las revistas de ciencias naturales se encuentra, a menudo, datos experimentales que se los procesa mediante el análisis de varianza o de regresión sin antes haber establecido si se conforman con la distribución normal. Mediante este procedimiento se invalida la realización de las pruebas de significación y se derivan conclusiones equivocadas.

2.—Consideraciones fundamentales.

Con estos antecedentes, vale revisar los aspectos relevantes de la teoría con el fin de establecer los criterios que deben aplicarse en el manejo de los datos experimentales.

Por razones obvias, presentamos a continuación los conceptos fundamentales acerca de las distribuciones normal, logarítmico-normal, binómica y de Poisson con el fin de establecer las relacio-

nes que existen entre ellas y las transformaciones que deben aplicarse para hacer viable la realización del análisis de varianza.

a) Distribución normal.— Esta distribución se caracteriza por ser simétrica a causa de que la probabilidad de que los desvíos de la media $(x - \mu)$ sean positivos o negativos es igual y que estos desvíos acontecen al azar y tienden a ser ínfimos.

La función de distribución $p\{x\}$ se la representa mediante la función típica $\Phi(u)$, una vez que

$$p\{x\} = \frac{1}{\sigma} \Phi(u) \text{ donde } u = \frac{x - \mu}{\sigma}$$

siendo x la variable aleatoria, μ la media y σ la desviación típica.

Actualmente, $p\{x\}$ representa la probabilidad de la realización de un evento que se conforma con la distribución normal:

$$p\{x\} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad (-\infty < x < \infty)$$

En esta ecuación, μ y σ^2 representan la media y la varianza que de hecho son los parámetros de la distribución.

La función de distribución cumulativa de x_1 , se define mediante la ecuación:

$$p\{x\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-(x-\mu)^2/2\sigma^2} dt$$

La distribución normal es característica de los errores de medida, experimentales, o aleatorios. Un ejemplo típico de datos que se conforman con la distribución normal constituyen los incrementos de peso de cobayos mediante el suministro de concentrados de composición diferente.

TABLA 1.—INCREMENTOS DE LOS PESOS (GRAMOS) DE COBAYOS. ALIMENTADOS CON CUATRO DIETAS.

Replicaciones	Concentrados			
	A	B	C	D
1	24,8	25,0	30,4	26,3
2	27,7	28,3	26,6	22,6
3	28,5	25,0	28,5	25,5
$\bar{x} =$	27,0	26,1	28,5	24,8
$S^2 =$	3,79	3,73	3,61	3,79

Tal como queda demostrado, la distribución normal se caracteriza porque las medias son independientes de las varianzas y a su vez éstas son homogéneas.

b) Distribución logarítmico-normal.— Se afirma que una variable aleatoria se conforma con la distribución logarítmico-normal cuando los logaritmos de la variable se conforman con la distribución normal.

La función de distribución acumulativa se define a partir de la ecuación:

$$p(x) = \Phi(u)$$

donde

En la realidad, se observa que la distribución de los tamaños de los pesos de las partículas de un polvo disperso se conforman con la distribución logarítmico-normal.

c) Distribución binómica.— La distribución binómica es característica de atributos duales: machos versus hembras, muertos y vivos, pequeños y grandes, verdes y maduros, etc.

Para distinguir las frecuencias de estas dos categorías se asignan las probabilidades p y q , reconociendo que $p + q = 1$. Además, si k representa el número de observaciones por muestra, se establece que la función de distribución es:

$$P(x) = \begin{bmatrix} k \\ x \end{bmatrix} p^x q^{k-x} \quad (x \geq 0)$$

En esta ecuación, x es la variable aleatoria, siendo k el parámetro discreto y p el continuo, una vez que varía de 0 a 1.

$$u = \frac{\log x - \log \mu}{\sigma} \quad (0 < x < \infty)$$

En esta ecuación, la media μ no presenta la media de la variable x , sino que μ se define mediante $\log \mu$ que es la media de $\log x$; o sea que

$$m \{ \log x \} = \log \mu$$

$$m \{ (\log x - \log \mu)^2 \} = \sigma^2$$

Diferenciando la función de distribución acumulativa, se tiene:

$$p \{ x \} dx = \phi(u) du$$

o sea que

$$p \{ x \} dx = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(\log x - \log \mu)^2}{2\sigma^2}} d \log x$$

lo cual hace que

$$P \{ x \} = \frac{M}{x \sigma} \Phi(\mu)$$

$$\text{La media de } x \text{ es: } m \{ x \} = \mu \times 10^{(\sigma^2/2M)}$$

$$\text{o sea que } \log x = \log \mu + 1,1513\sigma^2.$$

Los parámetros característicos de la distribución binómica son la media, $\mu = kp$ y la varianza, $\sigma^2 = kp(1-p)$. En cambio, cuando se expresan estos parámetros porcentualmente, $\mu = p$ y $\sigma^2 = pq/k$.

Un ejemplo típico representa los datos obtenidos en un experimento realizado para determinar la letalidad de una nueva droga con cobayos en grupos de cuatro. La probabilidad de supervivencia por cobayo es 0,5, siendo por supuesto evidente que pueden sobrevivir o morir de 0 a 4 cobayos por grupo. En total, se tratan 96 cobayos, siendo los resultados los siguientes:

$$\text{donde } M = \log e = 0,4343$$

TABLA 2.—SUPERVIVENCIA DE COBAYOS EN GRUPOS DE CUATRO INYECTADOS CON UNA DROGA.

Cobayos Sobrevivientes	Frecuencias de Supervivencia	
	Experimental (f_i)	Teórica ($f_{t,i}$)
0	4	6
1	20	24
2	40	36
3	25	24
4	7	6
	96	96

Se aprecia que para esta muestra, la medida y la varianza son:

$\bar{x} = 2,11$ y $s^2 = 0,92$, respectivamente.

Las características de esta distribución binómica se definen por el hecho de que cada grupo contiene 4 cobayos y que la probabilidad de supervivencia de cada cobayo es 0,5. Con este antecedente, las frecuencias teóricas $f_{t,i}$ se dan en la tercera columna de la Tabla 2.

Los parámetros de esta población binómica son $\mu = 2$ y $\sigma^2 = 1$. Por otro lado, se establecen estos parámetros directamente como sigue:

$$\mu = kp = 4 \times 0,5 = 2$$

$$\sigma = (kpq)^{1/2} = (4 \times 0,5 \times 0,5)^{1/2} = 1.$$

En el caso de expresar estos parámetros porcentualmente, se tiene:

$$\mu = p = 0,5 \text{ o sea } 50\% \text{ y } \sigma = (pq/k)^{1/2} = (0,5 \times 0,5/4)^{1/2} = 0,25 \text{ o sea } 25\%.$$

Se advierte que p no es solamente la probabilidad de un evento, sino que al mismo tiempo es la media μ de la población que representa el porcentaje de eventos "sobrevivencia o mortalidad"

y que σ es la desviación típica de la población. Finalmente, vale advertir que cuando los datos experimentales se conforman con la distribución binómica, la media está relacionada a la varianza, pues tal como queda puntualizado, cuando $p = q$ de hecho, $\sigma = 0,5 \mu$.

Otro ejemplo de datos que se conforman con la distribución binómica representan los obtenidos en un experimento para investigar la efectividad de cuatro drogas: A, B, C y D en pacientes divididos en grupos de 100 c/u.

TABLA 3.—EFECTO ANALGESICO DE CUATRO DROGAS SOBRE GRUPOS DE CIENTO PACIENTES.

Efecto	A	B	C	D
Positivo:	73	88	77	81
Nulo :	27	12	23	19
$\mu =$	0,27	0,12	0,23	0,19
$s^2 =$	0,197	0,106	0,177	0,154

d) Distribución de Poisson.— Los datos que se conforman con esta distribución son de carácter discontinuo, caracterizados por la media $\mu = p$ que tiene valor ínfimo. Por sus peculiaridades y aplicaciones, esta distribución se denomina también la distribución de eventos raros.

En la distribución binómica, cuando p varía de modo inverso a k , de manera que a medida que $p \rightarrow 0$, $k \rightarrow \infty$, la distribución binómica tiende hacia la de Poisson.

Partiendo de la distribución binómica, se llega a la de Poisson que se define mediante:

$$P(x) = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, 2, 3 \dots$$

Se establece esta ecuación mediante la siguiente transformación:

$$p\{x\} = \binom{k}{x} p q^{k-x} = \frac{k(k-1)\dots(k-x+1)}{x!} \frac{\mu^x}{k^x} \left(1 - \frac{\mu}{k}\right)^{k-x}$$

Se conoce que la otra forma de esta ecuación es:

$$p\{x\} = \left[1 - \frac{1}{k}\right] \dots \left[1 - \frac{x-1}{k}\right] \left[1 - \frac{\mu}{k}\right]^{-x} \frac{\mu^x}{x!} \left[1 - \frac{\mu}{k}\right]^k$$

se aprecia además que a medida que $k \rightarrow \infty$:

$$p\{x\} \rightarrow \frac{\mu^x}{x!} e^{-\mu}$$

Igualmente, los términos entre paréntesis con corchetes cuadrados convergen hacia 1 de tal modo que el producto converge también hacia 1, una vez que el número de factores es finito,

mientras que $\left[1 - \frac{\mu}{k}\right]^k \rightarrow e^{-\mu}$. Así se

establece que la media de la distribución binómica, $\mu = kp$, es fija y que la varianza es $kp(1-p) = \mu(1-p) \rightarrow \mu$ lo cual es cierto cuando $k \rightarrow \infty$ y $p \rightarrow 0$. Por lo tanto, la distribución de Poisson se caracteriza por tener un solo parámetro, una vez que la media es igual a la varianza.

Se ilustra este hecho mediante el siguiente ejemplo. Supongamos que para una distribución, $p = 0,001$ y $k = 1000$. La media y la varianza de la distribución binómica que tiende hacia la de Poisson son: $\mu = kp = 1$ y $\sigma^2 = kp(1-p) = 1 - 0,001 = 0,999$ lo cual confirma lo que queda puntualizado.

Vale anotar que se conforman con la distribución de Poisson el número de colonias bacterianas en placas del mismo tamaño, o el número de plantas enfer-

mas en lotes de terreno de la misma área. Específicamente se dan a continuación datos de la supervivencia de las moscas domésticas en grupos de 100 tratadas por triplicado con 3 insecticidas.

TABLA 4.—SUPERVIVENCIA DE MOSCAS TRATADAS CON TRES INSECTICIDAS.

Replicaciones	Insecticidas		
	A	B	C
R ₁	15	15	12
R ₂	24	11	16
R ₃	18	19	20
$\bar{x} =$	19	15	16
$s^2 =$	21	16	16

Se aprecia que para estos datos, las medias y las varianzas tienden a ser iguales.

e) Conversión de la distribución binómica en normal.—La distribución binómica tiende hacia la normal a medida que k aumenta y p disminuye. Por ejemplo, para $k = 100$ y $p = 0,1$, la aproximación es satisfactoria, una vez que la asimetría de la curva de la distribución binómica se reduce a un mínimo de tal manera que se identifica con la distribución normal. Pues:

$$p\{x\} = \binom{k}{x} p^x q^{k-x} \cong \frac{1}{\sqrt{2\pi kpq}} e^{-\frac{(x - kp)^2}{2kpq}}$$

Se percibe que cuando k es un valor grande, x se conforma con la distribución normal con parámetros: $\mu = kp$ y $\sigma^2 = kp(1-p)$. Así se establece que la función típica u se define mediante:

$$u = \frac{x - kp}{\sqrt{kp(1-p)}}$$

que tiene la propiedad de conformarse aproximadamente con la distribución normal. Si se dividen el numerador y el denominador por k resulta que

$$u = \frac{h - p}{\sqrt{p(1-p)/k}}$$

de donde se deduce que $h = x/k$ se halla distribuido normalmente con la media p y la varianza $p(1-p)/k$.

3.—Transformaciones.

a) Introducción.—Uno de los métodos más útiles es el análisis de varianza que rinde resultados válidos cuando se

$$\frac{1}{C} (\nu \ln s_1^2 - \sum \nu_i \ln s_i^2) \quad \text{donde}$$

lo efectúa con datos que se conforman con la distribución normal.

Los datos experimentales que se obtienen se desvían, a menudo, de la distribución normal, razón por la que se vuelve imprescindible aplicar transformaciones que, sin alterar la variación característica de los datos, haga posible la realización correcta del análisis estadístico. El objetivo que se persigue es el de lograr la homogenización de las varianzas.

Se entiende claramente que lo que

interesa es establecer si las varianzas de las muestras, s_i^2 tomadas independientemente son estimaciones de la varianza, σ^2 de la población.

Justamente, si el número de muestras es k , es evidente que a cada una corresponden $(k-1) = \nu_i$ grados de libertad. Para determinar si en realidad las

varianzas s_i^2 de las muestras son estimaciones de σ^2 , se parte de la hipótesis de comparación:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2$$

A base de esta hipótesis que afirma que al ser las varianzas iguales, de hecho las muestras corresponden a la misma población, la estimación s^2 de la población es:

$$s^2 = \frac{1}{\nu} \sum \nu_i s_i^2$$

donde $\nu = \sum \nu_i$. En vista de que la esperanza matemática de s_i^2 es la varianza de la población σ^2 , o sea que:

$$E(s_i^2) = \sigma^2$$

se ha establecido^a que la expresión

$$C = 1 + \frac{1}{3(k-1)} \left(\sum \frac{1}{\nu_i} - \frac{1}{\nu} \right)$$

se conforma con la distribución χ^2 , chi-cuadrado, con $(k-1)$ grados de libertad. La prueba chi-cuadrado^b sirve para establecer si las varianzas s_i^2 son estimaciones de σ^2 lo que en el caso de confirmar la hipótesis de comparación, significa que las muestras tomadas independientemente corresponden a la misma población.

La aplicación de una transformación a un limitado número de observaciones es difícil porque no es fácil decidir

(a) M. S. Bartlett, Proc. Roy. Soc., A., **160**, 268 (1937).

(b) Luis A. Romo S., Métodos de Experimentación Científica, Edit. Casals, Barcelona (1964), p. 159 — 160.

acerca de las características de las mismas.

Una prueba preliminar consiste en agrupar los datos para determinar si las medias son independientes de las correspondientes varianzas. Se realiza esta prueba con el fin de determinar si se aplica el teorema de la tendencia central que afirma, en términos generales, que "a medida que aumenta el tamaño de la muestra, la distribución de los promedios de las muestras con el mismo número de datos tomados de la misma población, tienden hacia la distribución normal bajo la condición de que la varianza de la población sea finita". En caso de que la prueba sea negativa, se debe aplicar una transformación como la condición previa imprescindible para realizar el análisis estadístico de los datos experimentales.

Si consideramos que $u = f(x)$ donde x es la variable aleatoria, la función de distribución acumulativa se define mediante la ecuación:

$$P\{x\} = f(x)$$

y por lo tanto:

$$dP\{x\} = P\{x\} dx = \phi(f(x)) df(x)$$

y que

$$P\{x\} = \frac{1}{\sqrt{2\pi}} e^{-[f(x)]^2/2} f'(x)$$

lo cual indica que mientras x se conforma con la distribución asimétrica, la función $f(x)$ se conforma con la distribución normal.

Se aprecia que, en principio, siempre es posible transformar los valores de x en unos que correspondan a la distribución normal, una vez que existe la función.

$$f(x) = \frac{g(x) - g(\mu)}{\sigma}$$

En esta ecuación, $g(x)$ no incluye ningún parámetro desconocido y tiene la propiedad de sujetarse a la distribución normal caracterizada por la media $g(\mu)$ y la varianza σ^2 .

b) Formas de las transformaciones.—

Entre las transformaciones que deben utilizarse para convertir los valores de la variable aleatoria x en datos que se sujetan a la distribución normal, se tienen las siguientes:

1) Las transformaciones $g(x) = \sqrt{x}$ y $g(x) = \sqrt{x + \alpha}$ se aplican cuando la variable aleatoria x se conforma con la distribución Poisson.

La transformación $g(x) = \sqrt{x}$ tiene la singular ventaja de que no contiene parámetros que deben ser estimados a partir de la muestra y que la distribución resultante se caracteriza mediante dos parámetros: la media y la varianza de la variable transformada.

En cambio, la transformación $g(x) = \sqrt{x + \alpha}$ incluye en la distribución el parámetro adicional que debe ser evaluado o introducido a priori arbitrariamente.

Se aplica la primera forma cuando el número de observaciones que se subordinan a la distribución de Poisson por grupo es igual a $k > 10$. En el otro caso, cuando $k < 10$, se aplica la segunda forma, fijando el valor de $\alpha = 0,5$.

Para aplicar estas transformaciones se deben hacer cálculos preliminares con los datos experimentales obtenidos con el

fin de establecer si $\bar{x} \approx \sigma^2$, tal como se indica mediante los datos de la Tabla 4.

Los datos que se consignan en la Tabla 5 representan la transformación de los datos de la Tabla 4, de Poisson, en normales.

TABLA 5.—TRANSFORMACION DE DATOS QUE SE CONFORMAN CON LA DISTRIBUCION DE POISSON EN NORMALES.

Replicaciones	Insecticidas		
	A	B	C
R ₁	3,87	3,87	3,46
R ₂	4,90	3,32	4,00
R ₃	4,27	4,36	4,47
$\bar{x}$	4,37	3,85	3,98
s ²	0,27	0,27	0,26

Se aprecia que mediante esta transformación se logra homogenizar las varianzas y establecer que las medias de los tres grupos sean independientes de las varianzas.

2) Transformación de datos que se conforman con la distribución binómica.— Cuando se expresan los datos experimentales como frecuencias, partiendo de la relación entre $\mu = p$ y $\sigma^2 = p(1-p)/k$, para la frecuencia $h = x/k$ se puede determinar la función $y = g(k)$ de tal modo que la variable transformada y sea independiente de P.

La función de transformación es:

$$y = 2 \arccos \sqrt{h}$$

La varianza de la distribución es $\sigma^2 = 1/k$; pues, si h concuerda con la distribución normal con media P y varianza $\sigma^2 = p(1-p)/k$, resulta que $y = \arccos \sqrt{h}$ se conforma también con la distribución normal con la media $\eta = 2 \arccos \sqrt{p}$ y la varianza $\sigma^2 = 1/k$.

En cambio, cuando los datos experimentales que se conforman con la distribución binómica se expresan en porcentajes, la transformación apropiada es la angular que viene dada por

$$\theta = \arccos \sqrt{p}$$

El análisis de varianza o de regresión se efectúa con los valores de θ que se encuentran tabulados (a).

A continuación, nos remitimos a los datos obtenidos en experimentos por triplicado para determinar el porcentaje de germinación de semillas de cuatro variedades de una hortaliza.

TABLA 6.—PORCENTAJES DE SEMILLAS NO GERMINADAS DE CUATRO VARIEDADES DE UNA HORTALIZA.

Replicaciones	Variedades			
	A	B	C	D
I	13	18	17	13
II	8	12	22	18
III	9	15	24	12
$\bar{x}$	10,0	15,0	21,0	14,3
s ²	7,0	9,0	13,0	10,3

Por las razones que quedan puntualizadas, para realizar el análisis de varianza, estos datos deben ser transformados para operar con los correspondientes valores de θ .

TABLA 7.—VALORES DE θ DE LOS DATOS CONSIGNADOS EN LA TABLA 6.

Variedades de una hortaliza				
	A	B	C	D
	21,1	25,1	24,4	21,1
	16,4	20,3	27,3	25,1
	17,5	22,8	29,3	20,3
$\bar{x}$	18,3	22,7	27,0	22,2
s ²	6,04	5,76	6,07	6,04

- a) R. A. Fisher y F. Yates, Statistical Tables for Biological, Agricultural and Medical Research, Oliver & Boyd Ltda., Edinburgh (1954).

3) Transformación logarítmica.—Existen problemas de investigación en los que un análisis preliminar de los datos tabulados indica que las medias son proporcionales a las varianzas. Cuando esto sucede, la teoría enseña que los datos corresponde a una distribución asimétrica con desvío hacia la izquierda.

La ecuación de la transformación es:

$$g(x) = \log(x)$$

que indica que los logaritmos de la variable aleatoria, se conforman con la distribución normal.

Se ilustra la aplicación de la transformación logarítmica mediante los datos obtenidos en un experimento para determinar el número de moscas en grupos de 10 c/u que se paralizan por el efecto de un insecticida sistémico después de 30, 60 y 90 segundos del tratamiento.

TABLA 8.—EFECTOS DE UN INSECTICIDA SISTEMICO SOBRE LOS TIEMPOS REQUERIDOS PARA PRODUCIR PARALISIS EN MOSCAS.

Replicaciones	Tiempos Transcurridos		
	30 seg.	60 seg.	90 seg.
1	2	5	6
2	5	3	6
3	2	2	9
4	3	6	3
$\bar{x}$	3	4	6
s^2	2,0	3,3	6,0

Para realizar con estos datos el análisis de varianza con el fin de determinar si en realidad las cuentas obtenidas de las moscas paralizadas en los tres tiempos de observación son diferentes, se realiza la transformación logarítmica, tal como se muestra en la tabla 9.

TABLA 9.—TRANSFORMACION LOGARITMICA DE LOS DATOS DE LA TABLA 8.

Replicaciones	Tiempos Transcurridos		
	30 seg.	60 seg.	90 seg.
1	0,301	0,700	0,778
2	0,700	0,477	0,778
3	0,301	0,301	0,954
4	0,477	0,778	0,477
$\bar{x}$	0,445	0,564	0,747
s^2	0,035	0,034	0,039

Tal como queda demostrado mediante esta transformación, se logra homogeneizar las varianzas sin menguar la variación de las frecuencias observadas que se manifiestan en las medias de las observaciones transformadas.

La Urgencia de una Renovación.

Aparte de las observaciones que quedan puntualizadas, se debe reconocer que el abuso de los métodos estadísticos por el desconocimiento de la teoría de las probabilidades, se patentiza en los modos que se utilizan para analizar erróneamente los datos de experimentos que se han realizado sin sujeción a un plan de investigación.

La teoría del planeamiento de experimentos científicos que ha alcanzado un alto grado de perfeccionamiento es desconocida en los países atrasados. Al respecto, cuán interesante sería que esta materia sea acogida y cultivada por los matemáticos.

En las instituciones de educación superior de algunos países latinoamericanos se enseña la Estadística a niveles que corresponden a la secundaria en los países desarrollados y en otras, simplemente se ignora que esta ciencia ha evolucionado enormemente desde su iniciación en los siglos XVIII y XIX.

El autor de esta contribución guarda la firme convicción de que el cultivo de la ciencia en los países subdesarrollados no podrá progresar jamás, mientras no se elimine el facilismo en todos los que-

haceres y el temor que existe por la incorporación, en los planes de estudios, de cursos de ciencias matemáticas y físicas a niveles acordes con la exigencia del mundo moderno. A menudo, se ignora que un objetivo fundamental del quehacer científico es el de alcanzar, a base de un limitado número de experiencias, generalizaciones que permitan resolver problemas y adquirir una más amplia y correcta interpretación del Universo.

Para corregir las gravísimas deficiencias de la enseñanza de la Estadística en las instituciones de educación superior, el prerequisite indispensable e insusti-

tuible debe ser un curso de la teoría de las probabilidades. Con este soporte teórico, la Estadística podrá ser la herramienta útil para el análisis correcto de los errores que vician los resultados de los experimentos, y la teoría del planeamiento de experimentos científicos, la base para la racional ejecución de la investigación.

Para terminar cabe referirnos a un pronunciamiento del Profesor R. A. Fisher^a que afirma que sin el conuoso franco de las Matemáticas es imposible entender el alcance de los métodos estadísticos.



ÁREA HISTÓRICA
DEL CENTRO DE INFORMACIÓN INTEGRAL

- a. Anotaciones del autor acerca del curso sobre la Filosofía de la Investigación Científica dictado por el Profesor R. A. Fisher, University of Cambridge (1955).

