Norte Grande, Inst. Geogr., Univ. Católica de Chile
Nº 6 (1978-1979)
Santiago, Chile

TEORIA Y APLICACION DE LA TRANSFORMACION DE VARIABLES EN GEOGRAFIA

ALDEN GAETE JENICEK (*)
Universidad Católica de Chile, Instituto de Geografía

SUMMARY

The validity of conclusions drawn from an Analysis of Variance depends upon the extent to which the assumptions of the technique are satisfied by the data. The assumptions are those of normality randomness and independence, additivity and homoscedasticity.

The present study contains three basic sections. The first demostrates that a situation where the assumptions are not met may be remedied by the creation of a new variable obtained by transforming the original data. In the second, the reasons which may cause the data to differ from the assumptions are discussed in conjuntion with descriptions of several appropriate transformations. Finally, the third section presents some theoretical and practical considerations arising from data transformation, also mentioning some non-parametric techniques which can be used instead of the usual transformations.

RESUMEN

La validez de las conclusiones provenientes de un Análisis de Varianza, depende de la relativa exactitud con que las observaciones satisfacen los supuestos implícitos en dicho análisis y que se refieren a la normalidad, la aleatoriedad e independencia, la aditividad y la homoscedasticidad.

El presente trabajo comprende tres secciones básicas. La primera de ellas muestra cómo el no cumplimiento de los supuestos puede evitarse mediante el uso de una nueva variable que se obtiene a través de la transformación de los datos originales. En la segunda sección, conjuntamente con estudiarse las causas que originan el no cumplimiento de los supuestos, se ve un conjunto selecto de transformaciones útiles en la investigación. Por último, la tercera sección presenta algunas consideraciones teóricas y prácticas concernientes a la obtención de las transformaciones, mencionando finalmente ciertas técnicas no paramétricas que se pueden utilizar en reemplazo.

1. Introducción.

La transformación de variables es una de las herramientas utilizadas en el Análisis de Varianza. El Análisis de Varianza proporciona una gran ayuda cuando se desea llegar a resultados e interpretaciones correctas, respecto a la diferencia significativa entre valores promedios calculados para cierta característica en estudio proveniente de distintas áreas o grupos.

En el Análisis de Varianza se plantean una serie de supuestos que no necesariamente se cumplen en todos los casos. Con el propósito de evitar los problemas que esto acarrea, Krumbein y Miller (1954) formularon una metodología donde sólo se refirieron al supuesto de Nor-

(*) Dirección del autor: Alden Gaete J., Instituto de Geografía, Universidad Católica de Chile, Casilla 114-D, Santiago, Chile.

malidad. La presente investigación, en cambio, proporcionará los elementos teóricos y prácticos para solucionar el no cumplimiento de los supuestos implícitos en el Análisis de Varianza en la medida que ello sea posible.

El esfuerzo que al respecto se haga, quedará ampliamente justificado al poder utilizar los conceptos que aquí se expongan, en el enorme campo de aplicación que dentro de la Geografía moderna representa el Análisis de Varianza, lo que se puede constatar al estudiar los trabajos realizados por Davis (1971) que examina fuentes de variación; Knoss (1962), que estudió los cambios de valores entre sectores urbanos y rurales; Murdie (1969), que examinó modelos zonales y concéntricos de variación espacial en la estructura social urbana, etc.

2. Problemas que plantea el no cumplimiento de los supuestos involucrados en el Análisis de Varianza.

2.1.—*Consecuencias del no cumplimiento de los supuestos.*

Para que un análisis de varianza esté plenamente respaldado por argumentos estadísticos que aseguren su rigurosidad metodológica, debe cumplir los supuestos usuales que utiliza y que son los de aditividad, aleatorización e independencia, normalidad y homoscedasticidad. De no verificarse algunos de ellos, será necesario efectuar modificaciones que permitan solucionar el problema o, en casos extremos, operar con otro modelo matemático cuando fuese factible, vale decir, cuando los datos no se ajusten al modelo teórico del Análisis de Varianza que originalmente se eligió.

En vista de que a menudo se observan en la práctica investigaciones en las cuales efectivamente no se cumplen en forma satisfactoria tales supuestos, se desarrollarán, a continuación, aquellas técnicas más utilizadas que permitan solucionar dichos inconvenientes.

Una de las formas de abordar el problema consiste en cambiar la unidad de medición de los datos a través de lo que se conoce como "transformación de variables". Mediante esta técnica se logran, la mayoría de las veces, resultados altamente satisfactorios, salvo en algunos casos en los cuales sólo se llega a soluciones aproximadas o definitivamente no recomendables. Otra forma alternativa que se plantea para el mismo problema, consiste en cambiar el modelo teórico originalmente propuesto. Por último, otro camino que se puede seguir es el de reemplazar las variables empleadas en el análisis que utiliza la estadística paramétrica por aquellas empleadas en la estadística de libre distribución. Es necesario señalar que entre ambos extremos existen soluciones intermedias a las ya enunciadas.

Dixon y Massey (1970) afirman: "Lograr el cambio específico en la escala de medición es, en general, difícil de determinar, dependiendo en gran medida de la experiencia del Investigador, el obtener la transformación precisa de las variables". Como norma, se requiere un gran número de observaciones que sirvan de antecedentes para la elección de cierta decisión referente al tipo de transformación más conveniente.

El problema se torna a veces complejo, al presentarse dentro de una misma investigación el no cumplimiento de dos o más supuestos en forma simultánea. Desde este punto de vista, la mera aplicación de determinada transformación no logra corregir el conjunto de alteraciones derivadas.

Estableciendo un orden de prioridades respecto al grado de importancia que significa el no cumplimiento de los supuestos, Snedecor (1964), establece: "La aditividad es un requisito esencial y la homoscedasticidad le sigue en importancia". También refiriéndose a la importancia de los supuestos, Bartlett (1947) afirma que "con la transformación de los datos en los cuales la componente residual tenía originalmente una distribución normal, se perderá parcialmente dicha propiedad después de la transformación". Sin embargo, esto no será muy serio según Bennet y Franklin (1954) dado que:

1. Una heteroscedasticidad sistemática frecuentemente viene acompañada por una asimetría en la distribución de los componentes aleatorios, la cual se corrige parcialmente por la transformación.
2. Pérdidas moderadas del requisito de normalidad son relativamente poco importantes en el análisis de varianza.
3. Las transformaciones que comúnmente se utilizan no significan pérdidas sustanciales en la eficiencia de los estimadores.

2.2.—*Requisitos que deben cumplir las Variables Transformadas.*

Las condiciones que se le exigen a la transformación de variables aplicadas dentro del Análisis de Varianza que plantea Federer (1955) son:

1. La varianza de las observaciones transformadas no debe ser afectada por cambios que se verifiquen en la media aritmética, vale decir, debe existir independencia entre ambos estadígrafos.
2. La variable transformada debe distribuirse normalmente.
3. Las observaciones ya transformadas deben poseer una media aritmética muestral que sea un estimador eficiente de la media aritmética poblacional, para cualquier grupo particular de mediciones.
4. Los datos transformados deben poseer la propiedad que los efectos que se investigan sean lineales y aditivos.

La combinación de dos o más de estas condiciones precedentemente señaladas permite a veces determinar las restantes. Por ejemplo, las condiciones 1, 2 y 4 determinan la condición 3. En otras ocasiones, la transformación que se aplique para satisfacer el requisito 1 posee, a menudo, la cualidad de aproximar la distribución a la normalidad, aparte de tener la ventaja de minimizar el sesgo producido por la dependencia entre la media aritmética y la varianza que se presentaría, según la distribución de las observaciones originales.

En todo caso, el aceptar completamente que el supuesto de normalidad se logre como efecto indirecto de la transformación que se aplica para obtener la independencia entre los dos estadígrafos mencionados, hay que tomarlo con reserva, pues en casos extremos si la aceptación de la normalidad va acompañada con un gran margen de error, puede conducir a invalidar la significación de los tests que se apliquen.

En aquellas ocasiones donde se posee la información suficiente respecto a la variabilidad de los datos, se recomienda abandonar las ventajas que se presentan en el simple análisis de varianza que exige el requisito planteado en el punto 1, y se prefiere utilizar una escala de medición que sea apropiada para satisfacer, en mayor medida, la condición en la cual evidentemente se ponderan las observaciones por sus respectivos valores y que están dependiendo de la variabilidad conocida. Esta metodología, tratada en detalle por Beall (1942), se aplicará más adelante en un caso concreto dentro de la transformación denominada de

"Probit" (Cfr. párrafo 4.1.3) y que se utiliza precisamente para proveer de una escala lineal, aplicable a variables que vienen expresadas en porcentajes.

Se debe agregar, por último, que cualquier investigación se simplificará si es que se opera con observaciones en las cuales los diversos efectos: tratamientos, bloques, etc., sean lineales y aditivos.

Es importante destacar un hecho que menciona Li (1969) y que debe tenerse presente en cualquier tipo de transformación de variables, el cual señala que el cambio debe hacerse para cada uno de los valores observados y luego trabajar como si las medidas originales fuesen nuevas unidades, y no comenzar a trabajar a partir de algún estadígrafo derivado al cual recién se le aplicará la transformación, tal como la media aritmética o la suma de cuadrados. De esta manera los cálculos subsiguientes que exigen el análisis de varianza, se basarán sobre esta nueva serie de datos.

Las relaciones que se hagan entre las variables en estudio estarán frecuentemente influenciadas según el tipo de escala con que se miden. De esta manera las interpretaciones de los resultados sólo serán válidas en relación con la escala particular que se empleó. Vale decir, un cambio de escala conduce a un cambio de relaciones entre las variables, y por consiguiente, a un cambio de la interpretación.

3. El problema de la relación de dependencia entre la media aritmética y la varianza.

Dentro de las condiciones exigidas que permiten asegurar una rigurosidad metodológica del proceso seguido en el análisis de varianza, se presenta la de efectuar un estudio acerca del comportamiento de los residuos, o, más concretamente, de la varianza del error. Ello tiene por objeto verificar si dicha varianza tiende a cambiar a medida que lo haga la media aritmética. De conocerse y cuantificar esa posible relación de dependencia, se podrá determinar el tipo de transformación específica que será necesario aplicar.

Numerosas investigaciones presentan cierto tipo de distribución en la cual existe una relación de proporcionalidad o dependencia entre la media aritmética y la varianza. A raíz de este fenómeno se han planteado diversos tipos de transformaciones que independizan ambos estadígrafos.

Dentro de las distribuciones más utilizadas en las investigaciones y que presentan tal relación de dependencia, se analizarán, entre otras, la BINOMIAL; LA POISSON Y LA EXPONENCIAL.

Para el caso de la Distribución Binomial, su varianza viene expresada por: $\sigma^2_x = n.p.q.$ En cambio, su media aritmética por: $m = n.p.$

Luego, se puede comprobar que existe una dependencia entre ambos estadígrafos proporcional a "q" veces la media aritmética "m". Tal relación, expresada matemáticamente en forma general, viene dada por la ecuación:

$$\sigma^2_x = f(m) \qquad 3.1$$

Presentando esquemáticamente tales relaciones según las diversas distribuciones, se tendrá:

Distribución Binomial $\left\{ \begin{array}{l} \sigma^2_x = n.p.q. \\ m = p.q. \end{array} \right\} \longrightarrow \sigma^2_x = q.m. = f(m) \qquad 3.2$

$$\text{Distribución Poisson}\left\{\begin{array}{l}\sigma^2_x = \lambda \\ m = \lambda\end{array}\right\} \longrightarrow \sigma^2_x = m = f(m) \qquad 3.3$$

$$\begin{array}{l}\text{Distribución}\\ \text{Exponencial}\end{array}\left\{\begin{array}{l}\sigma^2_x = 1/\lambda^2 \\ m = 1/\lambda\end{array}\right\} \longrightarrow \sigma^2_x = 1/\lambda \,.\, m = f(m) \qquad 3.4$$

4. Transformaciones de uso más frecuente.

4.1.—*Transformaciones de Variables en Distribuciones Discretas.*

4.1.1.—*La transformación Raíz Cuadrada.*

A menudo las observaciones que se logran de una investigación se expresan mediante una simple enumeración o conteo, en vez de estar medidas por una escala de intervalos. La información así representada suele encontrarse en aquellos experimentos donde la medición de la variable en estudio debe expresarse en porcentajes, los cuales, según ciertas características, tenderán a seguir una distribución de Poisson o una Binomial. Tal como ya se demostró en el capítulo anterior, ambas distribuciones presentan el problema de dependencia entre la media aritmética y la varianza, además de poseer efectos no aditivos.

Numerosas son las investigaciones que abarcan extensas áreas geográficas como son: grado de ataque en huertos frutales provocado por cierto virus y que fue detectado a través de la fotografía aérea mediante el uso de rayos infrarrojos; efecto que produjo cierto insecticida en el control de una plaga que afectaba la agricultura de una zona del país, etc. La información que se entrega, se expresa generalmente por datos que reflejan porcentajes o un conteo de la variable en estudio. Tales datos siguen, por lo general, las distribuciones precedentemente citadas.

Los modelos que operan dentro del análisis de varianza requieren como condición la aditividad de los efectos, aparte de exigir una distribución de los errores de tipo Normal, con media y varianza independientes, siendo esta última constante; o sea: $e \sim N\ (0, \sigma^2)$. Sin embargo, se constata que aquel tipo de investigación descrita anteriormente, viola el conjunto de condiciones mencionadas. Para solucionar tales problemas, a menudo resulta efectivo reemplazar cada medida u observación original por su raíz cuadrada y operar con los nuevos datos a lo largo de todo el proceso del análisis de varianza. Como consecuencia de tal transformación, tratada extensamente por Bartlett (1936), resulta que las varianzas de cada muestra (tratamiento) se homogenizan bastante entre sí, aparte que se logra un beneficio adicional en aquellas distribuciones sesgadas hacia la derecha o hacia la izquierda, pues, en tales casos, se "acorta la cola larga". Por último, y como efecto principal de la transformación, debe señalarse que ésta permite independizar la media aritmética de la varianza, incluso estabilizando dicho estadígrafo de dispersión.

La varianza muestral de los datos originales: S^2_x se puede calcular a partir de parcelas repetidas dentro de los bloques (caso de un diseño en bloques) o, en general, mediante un subconjunto de elementos que se someten a diferentes tratamientos. Luego se compara esta varianza con la media aritmética muestral: $\overline{x}$, proveniente también de dicho subconjunto. Si se comprueba la relación:

$$S^2_x \approx \overline{x} \text{ ó } S^2_x = k.\overline{x} \quad (1)$$

entonces será necesario aplicar la transformación Raíz Cuadrada.

4.1.1.1 *Algunos puntos por considerar en la transformación Raíz Cuadrada.*

1. Las nuevas variables transformadas "y", resultantes de la extracción de la raíz cuadrada a cada una de las variables originales "x", guardan entre sí la relación $y = \sqrt{x}$. La varianza de "y" será independiente de la media aritmética, fenómeno que no ocurría con las variables originales que seguían una distribución de Poisson cuando $x \geqslant 10$. Por tratarse precisamente de la distribución de Poisson, es que para una simple enumeración o contaje, x representa un estimador de la media m, de manera que se verifica la relación: $s^2_x = m \approx x$.

La operatoria que deberá seguirse para obtener la transformación adecuada se expresa a través de los siguientes pasos: se estima en términos generales la presencia de una relación de dependencia entre los estadígrafos varianza y media aritmética, la que se expresa por la ecuación $\sigma^2_x = f(m)$. Donde σ^2_x es la varianza de las observaciones originales y que está en función de la media aritmética m. Para cualquier función g (x), que representa la observación original transformada, la cual se escribirá abreviadamente con la letra "y", se puede calcular a su vez la respectiva varianza, la que, de acuerdo a la nomenclatura y cálculos matemáticos del presente estudio, se expresa a través de la ecuación:

$$V\,[g(x)] = \sigma^2_y \approx \left(\frac{dg}{dm}\right)^2 . \, f(m) \qquad 4.1$$

donde σ^2_y es constante, que es lo deseado; vale decir, será independiente de las variaciones que experimente m. (2)

Es necesario calcular el valor que toma $\left(\frac{dg}{dm}\right)^2$, expresión que depende, a su vez, de la función g que se utilice.

El último término de la ecuación, vale decir, f(m) constituye un dato, una información con la cual se cuenta. Su expresión dependerá de la distribución específica con que se esté operando. Para el caso específico de

(1) Que expresada en un gráfico, se manifiesta por una relación lineal con coeficiente angular K.

(2) La demostración matemática detallada que permite llegar a esta expresión se describe en el capítulo siguiente.

la distribución Poisson, el valor de f(m) se deduce de la igualdad ya mencionada $S^2_x = x$. En consecuencia, f(m) será igual a:

$$f(m) = S^2_x \approx x = m.$$

Por lo tanto, la información x que se posee como dato es, en el caso de esta distribución, igual a su varianza.

Para calcular σ^2_y a partir de la nueva variable transformada $y = \sqrt{x}$, el método de máxima verosimilitud determina que se debe elevar al cuadrado el cuociente de las derivadas de $\sqrt{x}$ respecto a x, y luego multiplicar por f(m). Dicho de otro modo:

$$\boxed{\left(\frac{d\sqrt{x}}{dx}\right)^2 \cdot f(m)} \qquad 4.2$$

Desarrollando la fórmula se tiene que $\dfrac{d\sqrt{x}}{dx} = \dfrac{1}{2\sqrt{x}}$ 4.3

Además, para esta distribución (Poisson) se tiene que: f(m) = x

Al elevar al cuadrado se obtiene:

$$\boxed{\sigma^2_y = \left(\frac{1}{2\sqrt{x}}\right)^2 \cdot x \; \frac{1}{4} = 0{,}25} \qquad 4.4$$

Varianza que expresada en forma general (cuando $S^2 = km$) será:

$$\boxed{\sigma^2_y = \frac{k}{4}} \qquad 4.5$$

2. Bliss y Calhoun (1954) plantean un segundo punto referido a la estabilización de la varianza para el caso de muestras pequeñas. En aquellos casos donde el número de datos observados, representados por la frecuencia, es pequeño (cuando se manejan de tres a diez observaciones solamente), la varianza del error que se obtiene de la transformación aplicada a datos provenientes de frecuencias o muestras grandes (o sea $y = \sqrt{x}$), es menos estable que aquella obtenida por la transformación $y = \sqrt{x + 0{,}4}$, la cual posee igual varianza, vale decir: $\sigma^2_y = \dfrac{1}{4}$.

Igualmente, en el caso de muestras pequeñas, muchos investigadores utilizan alternativamente la transformación $y = \sqrt{x + 1/2}$, expresión que también es útil aplicar en aquellas investigaciones donde el contaje de la información incluye los valores cero que también participan dentro de la codificación de los datos. En tales circunstancias, el agregar la constante 1/2 a todos y cada uno de los datos evita los problemas que se presentan al tener que calcular la derivada de cero.

En todo caso, tanto la transformación que se obtiene al extraer la raíz cuadrada a la variable original más la constante 1/2, como aquella que le suma 0,4, proporcionan una varianza suficientemente estable para grupos de observaciones constituidos por tres o más datos.

3. Es posible verificar la concordancia de las variables transformadas respecto a la distribución Poisson, mediante un cálculo de bondad de ajuste dado por la fórmula:

$$\boxed{\chi^2 \approx \frac{\text{Suma de los Cuadrados del Error}}{\sigma^2_y}} \qquad 4.6$$

donde los grados de libertad de la χ^2 serán los mismos que le correspondan a la suma de los cuadrados del error.

4. La significación de la diferencia entre las medias aritméticas de los tratamientos se determinará en unidades de la variable transformada: y. Una vez finalizados los cálculos pertinentes, se podrán reconvertir los valores a partir de las observaciones transformadas. Por ejemplo, se podrá obtener el valor de $\bar{x}$, al elevar al cuadrado $\bar{y}$ (donde $y = \sqrt{x}$ o bien $y = \sqrt{x + 0,4}$). Sin embargo, esta operación conduce a una subestimación del valor x = la media aritmética de las observaciones originales. Esto se corrige calculando:

$$\bar{x} \approx \bar{y}^2 + S^2_y \qquad 4.7$$

o bien

$$\bar{x} \approx \bar{y}^2 + S^2_y - 0,4 \qquad 4.8$$

fórmulas que se obtienen con sólo recordar la relación inicial, o sea:

$$y = \sqrt{x + 0,4} \quad \text{elevando al cuadrado} \qquad 4.9$$

$$y^2 = x + 0,4 \quad \text{aplicando sumatoria} \qquad 4.10$$

$$\Sigma y^2 = \Sigma x + 0,4 \, . n \quad \text{dividiendo por n} \qquad 4.11$$

$$\Sigma y^2/n = \Sigma x/n + 0,4 \quad \text{y dado que } S^2_y = \Sigma\, y^2/n - \bar{y}^2 \qquad 4.12$$

$$S^2_y + \bar{y}^2 = \bar{x} + 0,4 \quad \text{despejando finalmente} \qquad 4.13$$

$$\boxed{\bar{x} = S^2_y + \bar{y}^2 - 0,4} \qquad 4.14$$

4.1.2. *La Transformación Angular*

También conocida bajo el nombre de Transformación Arco-Seno, se aplica a variables expresadas en porcentajes o proporciones.

Un tipo de información muy corriente en investigaciones es aquel en el cual las anotaciones proporcionan el número de individuos o elementos que poseen una determinada característica dentro de un grupo mayor "n" de ellos. El caso más usual y sencillo se presenta cuando los diferentes grupos que se analizan poseen el mismo o sensiblemente igual tamaño de elementos. Bajo esta condición, que como se verá, simplifica bastante el cálculo, se pueden citar ejemplos concretos que se dan en la realidad, tales como: "proporción de superficie forestal que cubre un área" (problema tratado por Haggett (1967) y King (1969); resistencia según variedad de cierta especie vegetal ante el ataque de una plaga o de agentes químicos. Refiriéndose a este último ejemplo, los resultados que se obtendrán se expresarán a través del porcentaje de plantas infectadas. Al tener que cuantificar el grado de ataque, se presenta el problema de lo difícil que es medir la superficie o extensión dañada en relación a la superficie total de la planta o, en general, del elemento en estudio; debe agregarse a este tipo de dificultad el hecho de que una vez comenzada la acción destructora a partir de un foco, la velocidad del ataque aumenta rápidamente. Con ello, al querer comparar el grado de virulencia del agente o, visto desde otro ángulo, el grado de resistencia de la planta según las diversas variedades, se deberá, para ser equitativa la comparación, medir rápidamente la totalidad de ellas, en cuyo efecto la medición porcentual otorga una herramienta de fácil y rápido cálculo.

Evaluada la información a través de porcentajes o proporciones y en la medida que la variable por cuantificar no tenga una ocurrencia de probabilidad pequeña, la frecuencia de las observaciones tiende a distribuirse binomialmente.

Dado que la distribución Binomial, como ya se demostró, presenta al igual que la Poisson una relación de dependencia entre la media aritmética y la varianza, se aplica la Transformación Angular a cada elemento con el fin de independizar ambos estadígrafos, logrando además en forma simultánea el alargamiento de ambos extremos de la distribución Binomial, a la vez que comprime el centro de la misma, aproximando en consecuencia dicha distribución a una Normal. Sin embargo, tal como se verá más adelante, esta transformación no será recomendable para aquellos casos en que "n" o en general el denominador que se utilice en la obtención de las proporciones, no sea constante o sensiblemente constante.

Si se designa por "p" a la proporción o fracción de los "n" individuos que poseen la característica en estudio y por otra parte "q" = 1-p representa la proporción de los restantes que no la poseen, luego: p + q = 1. Relacionando tales proporciones con la expresión trigonométrica:

$$\operatorname{sen}^2 \theta + \cos^2 \theta = 1 \qquad 4.15$$

donde θ es un ángulo, se puede considerar: $p = \operatorname{sen}^2 \theta$ expresión de la cual se deducen otras dos, al despejar el valor de θ y que son las siguientes: $\theta = \operatorname{sen}^{-1}. \sqrt{p}$; $\theta = \operatorname{arc\,sen} \sqrt{p}$, relaciones éstas que la Transformación Angular utiliza para obtener de una proporción o fracción cierto ángulo. De ahí el término de Transformación Angular.

4.1.2.1. *Algunos Puntos por Considerar en la Transformación Angular.*

1. La nueva variable θ, lograda mediante la Transformación Angular de la variable original p, como ya se adelantó, se expresa por la relación: $p = sen^2\theta$, siendo $q = cos^2\theta$. Por otra parte se sabe que la varianza de la variable original es igual a: $V(p) = \frac{p.q}{n}$ y la media aritmética es: $\mu = p$.

Aplicando en este caso una metodología idéntica a la seguida en la Transformación de Raíz Cuadrada, cuya secuencia viene ordenadamente tratada en la obra de Bennet y Franklin (1954) se logrará que la información original venga expresada en función de una nueva variable. Efectivamente, partiendo de la relación fundamental

$$\boxed{f(\theta) = f(p) \cdot \left(\frac{dp}{d\theta}\right)^2} \qquad 4.16$$

$$\text{y sabiendo que: } f(p) = \frac{n}{p.q} = \frac{n}{sen^2\theta \cdot cos^2\theta} \qquad 4.17$$

$$\text{y que: } \frac{dp}{d\theta} = 2 \text{ sen } \theta \cdot \cos \theta \qquad 4.18$$

se reemplazan estos dos últimos valores en la primera ecuación y se despeja f(p), previa simplificación, quedando finalmente:

$$\boxed{f(p) = 4.n} \qquad 4.19$$

Tanto Kempthorne (1967) como otros investigadores, coinciden en señalar que mientras θ represente los ángulos a través de los cuales se expresan los arco senos, la varianza correspondiente a la nueva distribución (dentro del contexto de tamaño de muestra grande), vendrá dada por la fórmula:

$$\boxed{\sigma^2_\theta = \frac{820,7}{n}} \qquad 4.20$$

En cambio, cuando θ se mide en radianes en vez de grados, la varianza de la nueva distribución se expresa por:

$$\boxed{\sigma^2_\theta = \frac{1}{4.n}} \qquad 4.21$$

2. Aquellos Análisis de Varianza que trabajan con ángulos provenientes de muestras pequeñas ($n \leqslant 30$) se deberán evitar en lo posible. De no existir otra alternativa, aún en tales circunstancias se deberá preferir utilizar el cuadrado medio del error antes que la variante esperada.

3. Estabilización de la varianza para porcentajes observados de 0 y 100 por ciento.

Con el fin de evitar la discontinuidad al término de la escala, se puede aplicar un ajuste recomendable en aquellos casos en que aparezcan valores tales como un 0 ó un 100 por ciento. Tal ajuste consiste en computar el cero por ciento pero asignándole otro valor y que es la cifra 1/4. Para aquellos casos en que se observen valores de cien por ciento, se colocará en cambio la expresión n—1/4. Luego se procede a transformar en ángulos tales porcentajes ajustados.

4. La bondad de ajuste de la información transformada con la distribución Binomial se puede verificar a través de la prueba χ^2 para concordancia. Para tal efecto, dada la suma de cuadrados del error: S.C.E. y la varianza: $\sigma^2_y = \dfrac{820{,}7}{n}$, se calcula el valor de:

$$\boxed{\chi^2 = \frac{S.C.E.}{\sigma^2_y}} \qquad 4.22$$

con los grados de libertad del numerador.

5. Tal como se señaló anteriormente, la operatoria que se sigue para lograr la Transformación Angular consiste en calcular el valor de θ donde: $\theta =$ arc sen. $\sqrt{p}$, en que "p" representa una proporción. Los resultados de esta operación se pueden encontrar tabulados en cualquier texto pertinente.

Dentro de las relaciones trigonométricas, se tiene que: arc sen x = sen^{-1}x; o sea, la función arc sen es igual a la del seno inverso. Expresando dichas relaciones mediante un ejemplo numérico, se tiene: el valor del arc sen 0,431 podrá leerse en las tablas pertinentes con el valor 41.03°, vale decir, el ángulo cuyo seno es 0.431. Expresando lo anterior en forma abreviada se tiene:

	sen 41.03° = 0,431
luego	arc sen 0,431 = 41.03°
o bien	sen^{-1}0,431 = 41.03°

En resumen, se puede concluir que se dispone para el uso habitual, de tablas donde aparece el valor de θ expresando los grados correspondientes a los diversos valores de p de manera que en la práctica la transformación angular es tan sencilla como la logarítmica y la raíz cuadrada.

Se puede observar por último que los porcentajes y los ángulos equivalentes verifican la siguiente relación:

PORCENTAJES :	0	25	50	75	100
ANGULOS EN GRADOS :	0	30	45	60	90

luego, cuando dos porcentajes sumen 100, sus ángulos respectivos sumarán 90.

6. Resultados aceptables, productos de la Transformación Angular, serán difíciles de obtener en aquellas investigaciones donde las muestras o grupos son de tamaño distinto, debiendo recurrirse, en tales circuns-

tancias, a métodos aproximados. Además, para aquellos casos donde los porcentajes de la información original estén entre 30% y 70%, no será necesario aplicar la Transformación Arc Sen, como tampoco lo será cuando el tamaño de los grupos es grande, circunstancias bajo las cuales se puede efectuar el análisis directo de los números observados, prescindiendo así de la transformación.

7. Los promedios que se desea comparar deben, al finalizar la investigación, expresarse nuevamente en proporciones o porcentajes. Para ello se deben reconvertir los promedios que durante el Análisis de Varianza se expresaron en grados, mediante el uso de tablas que poseen el arc sen p como argumento y las proporciones como funciones. Vale decir, tablas de expresión:

$$p = (\text{sen}\ \theta)^2 = \text{sen}^2\ \theta \qquad 4.23$$

8. Dado que las transformaciones originan observaciones con distribución aproximadamente normal, se puede utilizar la expresión planteada en la obra de Dixon y Massey (1970) que es:

$$z = \sqrt{n+1} \qquad 2\ \text{arc sen}\ \sqrt{(x+1)/(n+1)} - 2\ \text{arc sen}\ \sqrt{p} \qquad 4.24$$

y que proporciona los límites de confianza aproximados para p ó para contrastar la hipótesis : $p = p_o$.

4.1.3. *La Transformación de Probit.*

También conocida como "Transformación en Probits", se utiliza al igual que la angular, con porcentajes, especialmente cuando se trabaja con datos de mortalidad.

De la distribución normal, se puede obtener el área bajo la curva hasta (o a partir de) un determinado punto de las abscisas. Es decir, se logra la frecuencia relativa hasta (o a partir de) un valor específico de la observación tipificada.

Pues bien, la Transformación de Probit no es nada más que la inversa de la tabla normal; da el valor de la observación tipificada para valores específicos del área de frecuencias relativas hasta (o desde) aquel punto. Se basa evidentemente, en la misma distribución Normal y puede obtenerse fácilmente por interpolación a partir de una tabla Normal detallada.

Puesto que la variable tipificada tiene media cero, con el fin de no trabajar con números negativos, se acostumbra agregarle el número cinco. Por lo tanto, para un área determinada bajo la curva Normal, el valor correspondiente que se toma es:

$$\boxed{\text{PROBIT} = 5 + \text{Observación Tipificada}} \qquad 4.25$$

La figura 1, que se muestra en seguida, da un ejemplo de la transformación. Cuando el porcentaje bajo la curva Normal es del 20%, la correspondiente observación tipificada es: —0,84 y el probit resultante será: 5,00 — 0,84 = 4,16. Por la simetría que caracteriza a la distribución de las observaciones tipificadas, los valores correspondientes al 20% y el 80% tendrán igual valor numérico, pero con signo opuesto. Por lo tanto, sus correspondientes valores de probit sumarán: 10; por ejemplo, 4,16 + 5,84 = 10,00. Cuando dos porcentajes cualesquiera suman 100%, sus correspondientes valores de probit sumarán 10.

Figura 1

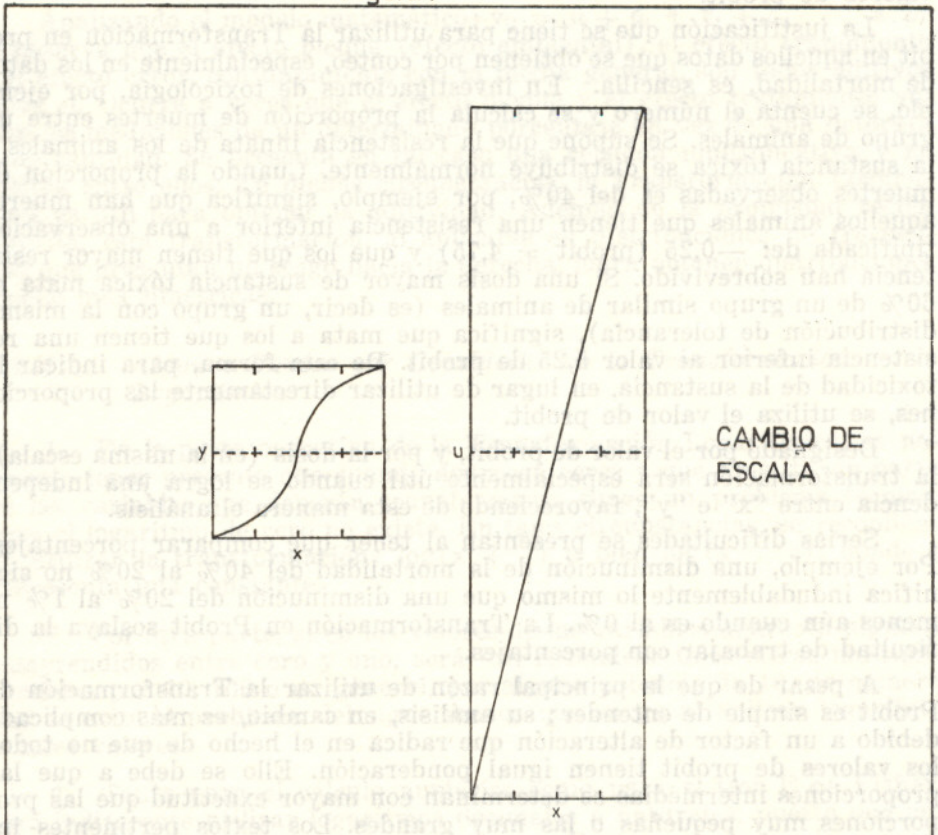
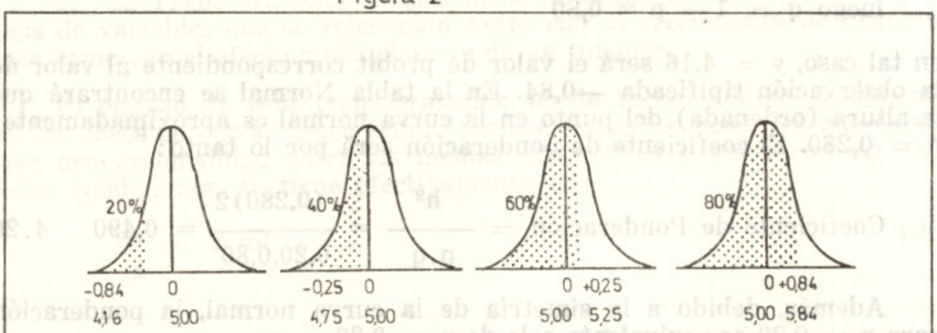


Figura 2



TRANSFORMACION DE PROBIT

A un determinado porcentaje del área, debajo de la curva normal, le corresponde una observación tipificada a la cual se le suma 5,00. Por lo tanto el probit es una variable normal con media 5,00 y varianza la unidad.

Con el fin de facilitar la aplicación de esta transformación, por lo general los textos especializados en esta materia agregan, a continuación de la representación gráfica del fenómeno, las tablas que contienen los valores de probit.

La justificación que se tiene para utilizar la Transformación en probit en aquellos datos que se obtienen por conteo, especialmente en los datos de mortalidad, es sencilla. En investigaciones de toxicología, por ejemplo, se cuenta el número y se calcula la proporción de muertes entre un grupo de animales. Se supone que la resistencia innata de los animales a la sustancia tóxica se distribuye normalmente. Cuando la proporción de muertes observadas es del 40%, por ejemplo, significa que han muerto aquellos animales que tienen una resistencia inferior a una observación tipificada de: —0,25 (probit = 4,75) y que los que tienen mayor resistencia han sobrevivido. Si una dosis mayor de sustancia tóxica mata al 60% de un grupo similar de animales (es decir, un grupo con la misma distribución de tolerancia), significa que mata a los que tienen una resistencia inferior al valor 5,25 de probit. De esta forma, para indicar la toxicidad de la sustancia, en lugar de utilizar directamente las proporciones, se utiliza el valor de probit.

Designado por el valor de probit, y por la dosis (en la misma escala), la transformación será especialmente útil cuando se logra una independencia entre "x" e "y", favoreciendo de esta manera el análisis.

Serias dificultades se presentan al tener que comparar porcentajes. Por ejemplo, una disminución de la mortalidad del 40% al 20% no significa indudablemente lo mismo que una disminución del 20% al 1% ni menos aún cuando es al 0%. La Transformación en Probit soslaya la dificultad de trabajar con porcentajes.

A pesar de que la principal razón de utilizar la Transformación de Probit es simple de entender; su análisis, en cambio, es más complicado debido a un factor de alteración que radica en el hecho de que no todos los valores de probit tienen igual ponderación. Ello se debe a que las proporciones intermedias se determinan con mayor exactitud que las proporciones muy pequeñas o las muy grandes. Los textos pertinentes incluyen una columna que indica los coeficientes de ponderación de los diversos porcentajes; estos coeficientes de ponderación se calculan de la siguiente manera:

Sea $p = 0{,}20$ la proporción observada
luego $q = 1 - p = 0{,}80$

en tal caso, $y = 4{,}16$ será el valor de probit correspondiente al valor de la observación tipificada —0,84. En la tabla Normal se encontrará que la altura (ordenada) del punto en la curva normal es aproximadamente: $h = 0{,}280$. El coeficiente de ponderación será por lo tanto:

$$\text{Coeficiente de Ponderación} = \frac{h^2}{p.q} = \frac{(0{,}280)2}{0{,}20.0{,}80} = 0{,}490 \qquad 4.26$$

Además, debido a la simetría de la curva normal, la ponderación para $p = 0{,}20$ es equivalente a la de $p = 0{,}80$.

4.2. *Transformación de variables en Distribuciones Continuas.*

4.2.1. *La Transformación Logarítmica.*

Constituye una de las transformaciones de mayor aplicación en la práctica de la Geografía. Se utiliza al manifestarse ciertos fenómenos a través de efectos que reflejan relaciones de proporcionalidad entre sí. Maxwell (1967) emplea frecuentemente esta transformación en sus investigaciones geomorfológicas.

Analizando el modelo matemático: $y_{ij} = \mu + \theta_i + \beta_j + \zeta_{ij}$ 4.27 se observa que el efecto bloque (efecto ambiental), el efecto tratamiento y el efecto error, conjuntamente con la media general, son todos aditivos entre sí. Ocurre a veces que el efecto bloque y el efecto tratamiento varían las mediciones de un determinado porcentaje en vez de una determinada cantidad. Se presenta en tal caso una relación multiplicativa y no aditiva entre ellos. En dichas circunstancias, la Transformación Logarítmica es útil para obtener la aditividad. Además de la propiedad recién descrita, esta transformación, al aplicarse en distribuciones de frecuencias que presentan un sesgo hacia la derecha, logra obtener distribuciones mucho más simétricas.

4.2.1.1. *Algunos Puntos por Considerar en la Transformación Logarítmica.*

1. En la parte operativa de la Transformación Logarítmica, es necesario tener presente que cuando aparecen ceros y que constituyen parte de las variables que deberán transformarse surge un problema técnico, pues el logaritmo de cero no existe. En tales circunstancias, se recomienda utilizar la transformación: $\log(x + 1)$ o bien $\log(x + c)$, donde c representa una constante.

2. Por otra parte, si en los valores de las variables aparecen algunos comprendidos entre cero y uno, será útil redondear tales cifras multiplicándolas por 10, 100 o en general, por alguna potencia de 10, de manera que se evite el problema del signo menos que surge al aplicar logaritmo a tales valores.

3. Tanto para el cálculo numérico como para el test F de Fisher, será indiferente utilizar logaritmos decimales o naturales, puesto que entre sí sólo difieren en un factor constante. En todo caso, por lo general se emplean logaritmos decimales, pues son más simples de aplicar, aparte del hecho que sólo se trabaja con los dos o tres primeros valores de la mantisa con el fin de facilitar el cálculo.

4. La Transformación Logarítmica se utiliza bastante para el análisis de variables que se relacionan tanto con el crecimiento de organismos como con el efecto multiplicativo de los mismos.

El empleo de esta transformación puede recomendarse igualmente en los casos en que se desea estabilizar las varianzas, cada vez que exista una proporcionalidad entre las medias aritméticas y las varianzas. En tales condiciones, se tiene efectivamente que:

$$\sigma^2_x = k . m^2_x \qquad 4.28$$

Luego, aplicando la misma metodología utilizada en las demás transformaciones, se tendrá:

$$\sigma^2_y = k . m^2_x \left[\left(\frac{d \ln x}{dx} \right) . \frac{1}{m_x} \right]^2 = k . m^2_x \left(\frac{1}{m_x} \right)^2 = k \qquad 4.29$$

5. Elección de la Transformación más Adecuada

5.1. *La Elección Teórica de una Transformación*

Además de los tres ejemplos recién expuestos, existen otras relaciones funcionales de dependencia que, por su frecuente aparición en problemas de aplicación práctica, será necesario detectar con el fin de individualizar ambos estadígrafos. Sin embargo, esto se logrará sólo conociendo a priori cuál es esa relación de dependencia específica y solamente entonces se podrá aplicar la transformación adecuada.

En tales circunstancias, la metodología por seguir consiste en comenzar expresando en forma general la relación de dependencia dada por la fórmula:

$$\sigma^2_x = f(m) \qquad 5.1$$

donde σ^2_x representa la varianza medida en la escala original y "m" es la media aritmética. Entonces, para cualquier función g(x) que representa la aplicación de cierta transformación a la variable original "x", ya sea la extracción de su raíz cuadrada, la aplicación de logaritmos, etc., se podrá demostrar que a esta nueva variable g(x) se le puede calcular su varianza cuya expresión será:

$$V\left[g(x)\right] \approx \frac{dg}{dm} \,.\, f(m) \qquad 5.2$$

Demostración: Siguiendo el planteamiento expuesto por Brownlee (1966), se tendrá:

Sea g(x) una función de la variable real "x".
Mediante la aplicación de las Series de Taylor, g(x) puede aproximarse a través de una función lineal, cerca de cualquier punto de m. Vale decir:

$$g(x) \approx g(m) + g'(m)\,(x-m) + g''(m) \,.\, \frac{(x-m)^2}{2} \qquad 5.3$$

De esta forma puede plantearse el siguiente problema:

Sea x una variable aleatoria con esperanza: $E(x) = m$. Entonces: $y = g(x)$ define a una nueva variable aleatoria en la cual:

$$y \approx g(m) + g'(m)\,(x-m) + g''(m) \,.\, \frac{(x-m)^2}{2} \qquad 5.4$$

Si x no varía demasiado alrededor de m, de tal manera que la varianza de x sea pequeña, se podrá en el momento de calcular las esperanzas a cada término, despreciar la expresión:

$E\ [(x-m)^2]$. De modo que la fórmula anterior queda:

$$E(y) \approx g(m) + g'(m)\,E(x-m) = g(m) \qquad 5.5$$

luego,

$$y - E(y) \approx g'(m)\,(x-m) \qquad 5.6$$

por lo tanto,

$$V(y) = E\,\{[y-E(y)]^2\} \approx g'(m)^2 \,.\, E\ [(x-m)]^2\ [g'(m)]^2 \,.\, V(x) \qquad 5.7$$

En general, V(x) dependerá de varios parámetros. En todo caso, se puede expresar en función del valor esperado m, o sea f(m), quedando entonces:

$$V(y) \quad [g'(m)]^2 \ . \ f(m) \qquad 5.8$$

Explicación de la demostración recién efectuada.

La demostración matemática parte suponiendo que la V [g(x)] es una constante; esto significa que será independiente ante cualquier variación de m. En tal caso, se reemplaza la expresión V g(x) por una C^2 y queda:

$$C^2 = \left[\frac{dg}{dm}\right]^2 . \ f(m) \qquad 5.9$$

De esta ecuación, se debe calcular el valor que toma $\left[\frac{dg}{dm}\right]^2$ y que depende de la función g que se use. El valor de f(m) se da como dato y dependerá de la distribución con que se trabaje. Finalmente, despejando se llega a:

$$\left[\frac{dg}{dm}\right]^2 = \frac{C^2}{f(m)} \qquad 5.10$$

donde extrayendo raíz cuadrada queda:

$$\boxed{\frac{dg}{dm} = \frac{C}{\sqrt{f(m)}}} \qquad 5.11$$

y si C^2 representa cierta constante: k^2, entonces:

$$\boxed{g = \int \frac{k.dm}{\sqrt{f(m)}}} \qquad 5.12$$

A modo de ejemplo, se desarrollarán algunos casos particulares, dejando simultáneamente establecido que dicha metodología es aplicable en otras circunstancias.

Primer caso particular

Se tiene la siguiente relación: $\sqrt{V \quad [g(x)]} = k_1.m$ (25) $\qquad$ 5.13

Para obtener f(m) se eleva al cuadrado la expresión anterior, resultando:

$$f(m) = k_2.m^2 \qquad 5.14$$

(25) Se denominará en general por k_1 a las constantes de proporcionalidad de m y cada vez que estén afectadas por alguna operación, se les irá cambiando la numeración del subíndice, comenzando por el número uno y agregándole una unidad por operación que se haga.

término que a su vez se reemplaza en la fórmula 5.11, con lo cual se obtiene:

$$\frac{dg}{dm} = \frac{C}{k_3 . m} \qquad 5.15$$

Dado que tanto C como k_3 son constantes, se puede llegar a la expresión:

$$\frac{dg}{dm} = k_4 \frac{1}{m} \qquad 5.16$$

Esta ecuación diferencial se resuelve despejando dg y luego integrando:

$$dg = k_4 . \frac{dm}{m} \qquad 5.17$$

$$\int dg = k_4 \int \frac{dm}{m} \qquad 5.18$$

$$\boxed{g = k_4 \ . \ \ln m} \qquad 5.19$$

En vista de que g representa la transformación que debe aplicarse a la variable original x con el objeto de independizar la varianza de la media aritmética, el resultado recién calculado indica claramente que para lograr la transformación correcta, a cada valor de la variable original se le debe calcular su logaritmo natural, debiéndose utilizar en adelante esta nueva variable con el fin de llegar a conclusiones más precisas en el análisis de varianza.

Segundo caso particular

A partir de la siguiente relación:

$$V \left[g(x) \right] = k_1 . m \qquad 5.20$$

Se tendrá directamente, en tales circunstancias, la igualdad:

$$f(m) = k_1 . m \qquad 5.21$$

por lo tanto

$$\frac{dg}{dm} = \frac{C}{\sqrt{k_1 . m}} = k_2 \frac{1}{\sqrt{m}} \qquad 5.22$$

despejando

$$dg = k_2 \frac{dm}{\sqrt{m}} \qquad 5.23$$

integrando

$$\int dg = k_2 \int \frac{dm}{m} \qquad 5.24$$

$$\boxed{g = k_3 \sqrt{m}} \qquad 5.25$$

Luego, en este caso particular en que la varianza guarda una relación de proporcionalidad directa con la media, Bartlett (1937) afirma que la transformación conveniente consiste en extraer la raíz cuadrada a las variables originales x, con lo cual se obtiene la nueva variable g que posee varianza constante y, por lo tanto, independiente de la media aritmética.

El método que permite llegar a la deducción de los dos casos particulares recién analizados señala concretamente en uno de sus puntos: "el valor de f(m) se da como dato y dependerá de la investigación que se esté efectuando".

5.2. *La Elección Empírica de una Transformación.*

A menudo no se dispone de ningún antecedente o justificación teórica (al menos en apariencia) que permita determinar a priori la utilización de cierta transformación. En tales casos conviene, entonces, efectuar una elección empírica.

Es posible orientar esta elección mediante la confección de un diagrama de dispersión cuya nube de puntos esté formada por las medias aritméticas y por las varianzas; de preferencia a escalas logarítmicas. Al ubicar las medias aritméticas en el eje de abscisas y las varianzas en el eje de las ordenadas, se presentan generalmente dos casos típicos:

1. Por una parte, la nube de puntos puede estar orientada paralelamente a la bisectriz del diagrama, lo que demuestra que la relación existente entre las medias y las varianzas es del tipo:

$$\log \sigma^2_x = \log m_x + \log k \qquad 5.26$$

o también

$$\sigma^2_x = k.m_x \qquad 5.27$$

En tales circunstancias, la transformación por utilizar será la Transformación Raíz Cuadrada.

Por otra parte, si la nube de puntos posee una inclinación dos veces superior a la bisectriz antes aludida, la relación media-varianzas será del tipo:

$$\log \sigma^2_x = 2 \log m_x + \log k \qquad 5.28$$

o también

$$\sigma^2_x = k m^2_x \qquad 5.29$$

En tales casos será la Transformación Logarítmica la que evidentemente se deberá aplicar.

Es, por lo tanto, la pendiente que posea la nube de puntos la que permitirá orientarse en la elección de una determinada transformación. Dicha pendiente puede estimarse por el coeficiente de regresión de los logaritmos de las varianzas en función de los logaritmos de las medias aritméticas o, más simplemente, por el coeficiente de la recta de mínimos cuadrados, que no es otra cosa que el cuociente entre la desviación típica de los logaritmos de las varianzas y la desviación típica de los logaritmos de las medias.

Cuando los términos son iguales se pueden reemplazar las medias y las varianzas, respectivamente, por las sumas y las sumas de cuadrados. Tal sustitución no modifica en nada la pendiente de la nube de puntos, en los casos de graficar a escalas logarítmicas.

2. Otras situaciones se pueden presentar evidentemente; según el caso, deberá elegirse la transformación más adecuada. En términos generales, para aquellos casos en que se tenga una relación cuyo coeficiente b proveniente de la relación entre el logaritmo de las varianzas y el logaritmo de las medias aritméticas, la transformación por utilizar (mientras que b sea diferente de 2) será:

$$\boxed{y = x^{1 - b/2}} \qquad 5.30$$

En tales condiciones, se tiene en efecto:

$$\log \sigma^2_x = b.\log m_x + \log k \qquad 5.31$$

$$\sigma^2_x = k.m^b \qquad 5.32$$

$$\sigma^2_y \approx k.m^b \left(\frac{dx^{/1-b2}}{dx} \right)^2_{m_x} = k(1 - b/2)^2 = k' \qquad 5.33$$

3. Otros investigadores han planteado métodos más rigurosos para obtener en forma empírica la elección óptima de una transformación. En la práctica, frecuentemente dichos métodos son, sin embargo, muy laboriosos, y no parece necesario insistir en ellos.

Independiente de las dificultades de cálculo, el principal inconveniente del cambio de variables es el de complicar la interpretación de los resultados. Dicha dificultad es de importancia relativamente secundaria en las Transformaciones Raíz Cuadrada y Logarítmica, pero será mucho más considerable en otras. Es por ello recomendable trabajar en lo posible con las transformaciones más simples.

6 Técnicas no Paramétricas utilizadas en la Transformación de Variables.

Al no proporcionar resultados satisfactorios ninguna de las transformaciones de variables expuestas en los capítulos anteriores, se puede recurrir a los métodos o herramientas empleadas en la estadística de libre distribución. Se denomina de "libre distribución", puesto que no está sujeta a ninguna distribución de probabilidades específica a diferencia de la estadística paramétrica. La hipótesis nula que se plantea en este caso no se circunscribe a ningún parámetro específico (como es el caso de la media aritmética que generalmente se emplea al trabajar con una distribución de probabilidad conocida). En el presente caso, sólo se trabaja con la distribución de las variables y se prescinde de los parámetros. De ahí proviene el término "métodos no paramétricos" con que se caracteriza a esta nueva herramienta.

Recientemente el análisis de varianza no paramétrico ha adquirido bastante popularidad debido a que es simple de computar y que permite mayor libertad de acción que el paramétrico al no ser tan restrictivo en sus exigencias, ya que no depende de los supuestos de normalidad ni de homoscedasticidad. Sin embargo, debe advertirse que de poder emplear satisfactoriamente el análisis de varianza paramétrico, los resultados que de él se obtengan tendrán mayor potencia y eficiencia que los provenientes de un análisis equivalente no paramétrico. (Conclusión que analiza en detalle Siegel, 1970).

Cuando el test se refiere solamente a la comparación de dos muestras (equivalente al test t de Student), se pueden emplear una serie de pruebas no paramétricas, entre las que sobresalen la Prueba U de Mann-Whittney y la Prueba de Wilcoxon para dos muestras.

Para comparar más de dos grupos o muestras en forma simultánea, el modelo equivalente al Diseño Completamente Aleatorizado empleado en la estadística paramétrica, es el que corresponde al Análisis de Varianza de una clasificación por rangos de Kruskal-Wallis. De igual manera, el Diseño en Bloques Completamente Aleatorizados posee su equivalente en la estadística de libre distribución y que es el Test de Fiedmann para Bloques Aleatorizados.

Ejercicios y aplicaciones concretas que se refieren a esta materia son tratados por Sokal y Rohlf y Siegel (1970), entre otros.

BIBLIOGRAFIA

BARTLETT, M. S., 1936.—"The Square-Root Transformation in the Analysis of Variance". *Suppl. Jour. Roy. Stat. Soc.*, 3, 38-78.

BARTLETT, M. S., 1937.—"Some examples of Statistical Methods of Research in Agriculture and applied Biology." *Suppl. Jour. Roy. Stat. Soc.*, 4, 137-183.

BARTLETT, M. S., 1947.—"The Use of Transformation". *Biometrics*, Vol. 3, 39 p.

BARTLETT, M. S. and KENDALL, D. G., 1946.—"The Statistical Analysis and the Logaritmic Transformation". *Suppl. Jour. Roy. Stat. Soc.*, 8, 128-138.

BEALL, G., 1942.—"The Transformation from entomological Field Experiments so that the Analysis of Variance becomes applicable". *Biometrika*, 32, 243-262.

BENNET, A. C. and FRANKLIN, L. N., 1954.—*Statistical Analysis in Chemistry and the Chemical Industry.* New York, John Wiley.

BLISS, I. C. and CALHOUN, W. D., 1954.—*An Outline of Biometry.* New Haven, Connecticut, Cooperative Corporation.

BROWNLEE, A. K., 1966.—*Statistical Theory and Methodology,* 2 ed. New York, John Wiley.

COCHRAN, W. G., 1947.—"Some Consecuences when the Assumptions for the Analysis of Variance are not satisfied". *Biometrics.* Vol. 3, pp. 22.

DAVIS, J. T., 1971.—"Sources of Variation in Housing Values in Washington, D.C." *Geographical Analysis.* 3:63-76.

DIXON, J. N. y MASSEY, J. F., 1970.—*Introducción al análisis estadístico.* 2 ed. México, Mc. Graw-Hill.

BEDERER, T. W., 1955.—*Experimental Design Theory and Application.* New York, The Mac. Millan.

FISHER, R. A., 1950.—*Statistical Methods, for Research Workers,* 11th, ed. Edimburg. Oliver and Boyd.

HAGGET, P., 1967.—*Locational Analysis in Human Geography,* 3 ed. London, Edward Arnold.

KEMPTHORNE, O., 1967.—*The Design and Analysis of Experiments.* 6 Ed. New York, John Wiley.

KING, J. L., 1969.—*Statistical Analysis in Geography,* Englewood Cliffs, N. Y. Prentice-Hall.

KNOSS, D. S., 1962.—*Distribution of Land Values in Topeka, Kansas.* University of Kansas, Bureau of Business and Economic Research.

KRUMBEIN, W. C. and MILLER, R. L., 1954.—"A Note on Transformation of Data for Analysis of Variance". *Journal of Geology,* Vol. 62, 1954, p. 192-193.

LI, C. C., 1969.—*Introducción a la estadística experimental,* Barcelona, Omega.

MAXWELL, C. J., 1967.—*Quantitative Geomorphology of some Mountains Chaparral Watersheds of Southern California.* DEPARTMENT OF GEOGRAPHY, North Western University, Evanston, Illinois.

MURDIE, R. A., 1969.—*Factorial Ecology of Metropolitan Toronto* 1951-1961, *An Essay on the Social Geography of the City.* (Research paper Nr. 116). Chicago: Department of Geography, University of Chicago, 212 p.

SIEGEL, S., 1970.—*Diseño experimental no paramétrico aplicado a las ciencias de la conducta.* México, Trillas.

SNEDECOR, W. G., 1964.—*Métodos estadísticos aplicados a la investigación agrícola y biológica.* 4 ed., México, Continental.