

Capítulo 6. Lenguajes estadísticos



Estadística es la rama de las Matemáticas que recoge, describe, cuantifica, tabula, interpreta los datos que se recogen de los hechos de cualquier naturaleza. Por su rigor, se ha convertido en el lenguaje preferido por los investigadores para exponer y publicar todo tipo de trabajo realizado.

La estadística puede ser de dos tipos:

■ Se llama estadística descriptiva a la que se hace cuando se recuenta la totalidad de una “población” y se condensan los resultados en datos globales y sintéticos, ya se haga de forma absoluta (numérica) o de forma relativa (porcentual o percentilada).

■ Se denomina estadística inferencial la que parte de una “muestra” para obtener unos datos “estadísticos”, cuyo valor será tanto más creíble cuanto la muestra sea más representativa de la población.

La más usual es la inferencia estadística, sobre

todo cuando las poblaciones son numerosas. Debido a ello, se requieren técnicas de muestreo científico, fórmulas para determinar la credibilidad o garantía de los resultados, cauces de expresión suficientemente claros y sólidos.

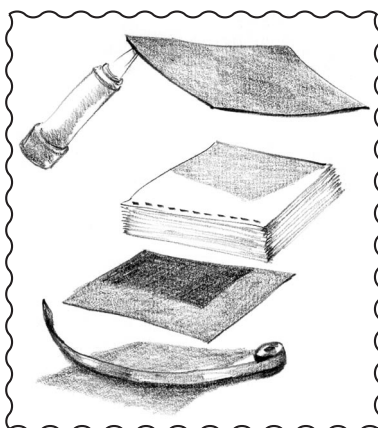
1. Valor de los lenguajes estadísticos

En los terrenos sociales relacionados con los valores religiosos o educativos se multiplican los usos estadísticos: cumplimientos, pertenencias, preferencias, aportaciones, etc.

Pero en lo relacionado estrictamente con la cultura, la psicología, la afectividad, la moralidad o la religiosidad, la Estadística encuentra barreras infranqueables: son las del misterio de la libertad humana en lo referente a los contenidos, la a diversidad y originalidad en relación a las personas, a la falta de alcance en lo referente a los sentimientos, a las ideas y a los criterios, a la subjetividad en cuanto afecta las interpretaciones.

Medida en psicología

La Psicología y la Pedagogía emplean las expresiones orales y escritas ordinarias para expresar los datos que descubren o investigan. Pero, también necesitan acudir a los lenguajes de las ciencias positivas para ganar en precisión, claridad y riqueza analítica. Tienen especial preferencia por





dos tipos de lenguajes: los matemáticos o estadísticos y los gráficos. Con todo, no los usan de forma diferente a como lo hacen las otras ciencias sociales o humanas, como puede ser la Sociología, la Economía o la Política.

Se precisan técnicas y formas especiales para expresar cuantitativa y cualitativamente los hechos psicológicos. Por ello, las técnicas y formas de medirlos con lenguajes y criterios matemáticos no siempre resulta fácil.

La psicometría recurre a las ciencias numéricas para medir. Y por eso estudia el valor y el uso de medidas absolutas (medias, medianas, modas, desviaciones medias, varianzas, desviaciones típicas, porcentajes) y de medidas relativas (puestos o rangos, niveles, percentiles, deciles, correlaciones).

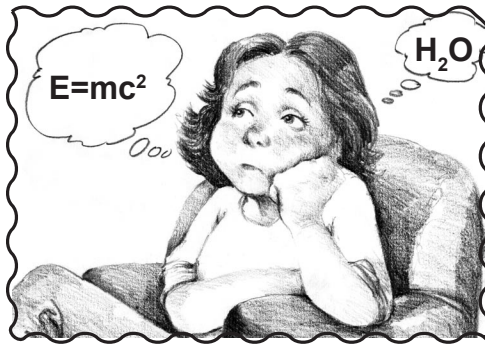
La Estadística y el cálculo de probabilidades, en la medida en que la geometría o uso significativo de figuras reclama la medida cuantitativa o cualitativa, es lo que más se emplea en la psicología.

Diversidad de hechos psicológicos y pedagógicos se expresan con lenguajes matemáticos: porcentajes de resultados, intensidad de intereses, fortaleza de las decisiones, autonomía de las actitudes, niveles de normalidad, predicción de valores, etc. No se puede investigar en estos terrenos sin entender algo los lenguajes psicométricos para situarse objetivamente ante la realidad de los educados y para valorar objetivamente su situación, no sólo escolar, sino personal.

La medida es un lenguaje imprescindible en la comunicación humana y por eso conviene tenerla también presente en la actividad educativa de todo tipo, incluidos

los terrenos religiosos, sociales y morales, en cuanto puedes ser susceptibles de medida.

La Estadística es la ciencia matemática que ayuda a justificar, usar y elaborar las principales medidas. Hay dos formas de usarla: siguiendo las líneas descriptivas de la estadística descriptiva, que es la que trata de datos procedentes de poblaciones; o prefiriendo la intuición o inducción que requiere la estadística inferencial, que es la que elabora datos procedentes de “muestras significativas”, tanto más fiables cuanto mejor reflejan fielmente la situación de la población en donde se obtienen.



2. Las medidas absolutas

Un recuerdo de las medidas de uso más frecuentes puede ayudar a entender su importancia y variedad. Las hay absolutas, que son las que recopilan y expresan datos globales; y las hay relativas, que implican referencia a variables interseccionadas.

Las medidas absolutas nos ofrecen datos recogidos con criterios o intenciones variadas. A unas se las llama medidas de **tendencia central**.

Son las siguientes:

■ **La MEDIA.** Es el valor medio de un grupo o serie de valores. Se representa por M y es la suma de los valores de cada sujeto del grupo, dividido el resultado por el número de sujetos del grupo.

Su fórmula es:

$$X \text{ ó } M = \frac{\text{Suma de } x}{\text{Número}} = \frac{Sx}{N^{\circ}}$$



■ **La MODA.** Es el valor más repetido en una serie o grupo. Se representa por Mo.

Su fórmula es:

$$Mo = a \dots b \dots c \dots [c] \dots c \dots d \dots e$$



■ **La MEDIANA.** Es el valor que ocupa el lugar central en una serie ordenada de valores. Se representa por Md.

Su fórmula es:

$$Md = a \dots b \dots c \dots [d] \dots e \dots f \dots g$$



Una serie es más armónica y homogénea cuanto más coinciden en ella la M, la Mo y la Md. Si discrepan, la serie o grupo se caracteriza por desproporción y por dispersión de sujetos extremos.

A otras medidas se las denomina de **tendencia dispersiva**.

■ **La DESVIACIÓN MEDIA.** Es el valor medio de las desviaciones de los sujetos con respecto a la media. Se representa por Vm o Dm.

Su fórmula es:

$$Dm = \frac{S d}{N} \quad \frac{\text{Suma de las Diferencias en relación a la Media}}{\text{Número de casos}}$$

■ **La VARIANZA.** Es la media de las desviaciones de los sujetos con respecto a la media de la serie, pero elevadas esas desviaciones al cuadrado. Se representa por una sigma (letra griega S) al cuadrado (2).

Su fórmula es:

$$Sg2 = \frac{S d^2}{N} \quad \frac{\text{Suma de las diferencias al cuadrado}}{\text{Número de casos}}$$



■ **La DESVIACIÓN TÍPICA o Sigma.** Es la raíz cuadrada de la varianza. Equivale, por tanto, a la raíz cuadrada de la suma de los cuadrados de las desviaciones individuales (desviaciones con referencia a la media).

Su símbolo es la Sigma. (S griega: Sg) .

Su fórmula se expresa así:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad \text{o también} \quad s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{N}}$$

O más vulgarmente dicho:



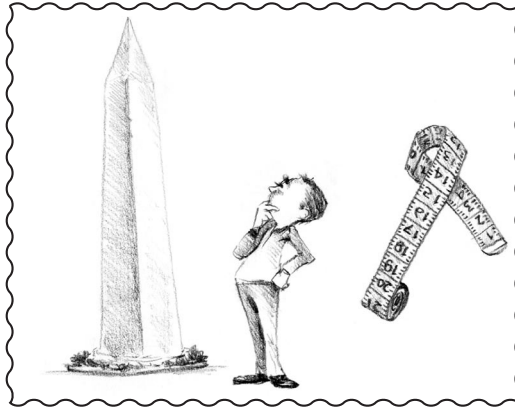
$$Sg : \sqrt{\frac{Sd^2}{N}} \quad \text{que se lee:}$$



Sigma o Desviación típica es la raíz cuadrada ($\sqrt{\quad}$) del cociente entre las sumas de las diferencias al cuadrado, dividido por el número de casos.

Media, Mediana y Moda son medidas de tendencia central. Significa que su referencia es el centro, la Media de una serie. Desviación media, Varianza y Desviación típica son medida dispersivas. Significa ello que su referencia es la dispersión referente a la media.

A ambas se las llama absolutas o cuantitativas, para que no se confundan con las cualitativas o relativas



3. Medidas relativas o cualitativas

Las medidas relativas son las que buscan una expresión numérica en función de una relación establecida con el grupo o serie. Las más usadas son los cuantiles y las correlaciones.

■ **Los CUANTILES** son los lugares que ocupan los sujetos de un grupo en relación a una serie idealmente ordenada. Esa serie puede dividirse en varios tramos o sectores: cuatro, diez, cien son los más usuales.

Se llaman CUARTILES, si la serie se divide en cuatro sectores. El símbolo es Q. Y la fórmula se expresa así:

$$Q_x = \frac{P^o_x \times 4}{N}$$



Se llaman DECILES, si la serie se divide en diez sectores. El símbolo es D. Y la fórmula es:

$$D_x = \frac{P^o_x \times 10}{N}$$



Se llaman CENTILES O PERCENTILES, si la serie se divide en 100 sectores o niveles. El símbolo es P. Y la fórmula es:

$$P_x = \frac{P^o_x \times 100}{N}$$



Usar los Percentiles (que no es lo mismo que porcentajes), es usar cuantiles que se refieran a escalas de cien espacios o escalones. Se emplean sobre temas y ámbitos pedagógicos ordinales (puestos, situaciones, resultados de pruebas objetivas y tests de conocimientos) y también psicológicos (baremos de tests y de escalas). Es decir el puesto que ocupa un sujeto en una escala ideal de 100 niveles. Si uno ocupa el Percentil 60, tiene 59 puestos

por debajo de él, es decir el 59% (o el 60%, según se mire) de los sujetos del grupo de referencia. Y tiene 39 puesto, el 39% (o el 40%, según se mire), por encima.

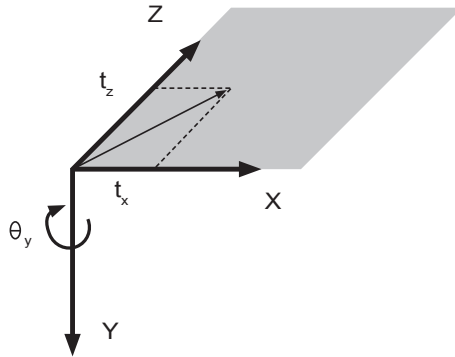
■ **Se usan también las CORRELACIONES.**

Correlación es la tendencia de dos fenómenos o rasgos a variar concomitantemente. Si dos fenómenos o rasgos tienen conexión o interdependencia entre sí, al variar uno varía el otro. Podemos establecer la intensidad cuantitativa de esta interrelación o variación concomitante y expresar el resultado de forma numérica, en referencia a la unidad o de otra forma.

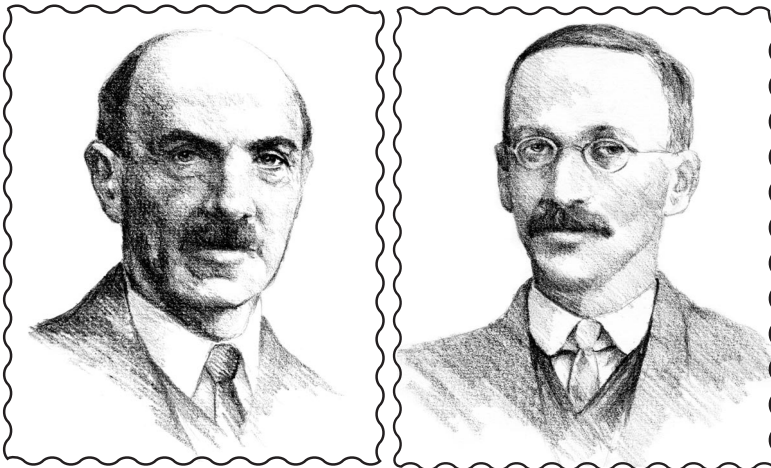
La expresión numérica se denomina coeficiente o índice de correlación. Los matemáticos han usado muchas fórmulas para medir y expresar la correlación. Casi todas ellas han tendido a usar la unidad como valor de referencia, llamando índice de correlación a la cuantía que va entre 0 y 1.

De modo que, si el índice es 0, la correlación es nula; si es 1, la correlación es perfecta o total. En la medida en que se halla en torno a 0,50, la correlación es media. Si el índice es positivo: + 1, + 0,70, + 0,50, + 0,30, indica que los dos fenómenos correlacionados varían en la misma dirección: al subir uno, el otro sube; y al bajar uno, el otro baja: Por ejemplo, los tiempos de estudio y las calificaciones escolares, la cantidad de comida y la obesidad.

Y si el índice sale negativo (- 1, - 0,70, - 0,50, - 0,30), expresa una correlación inversa. Al subir un factor baja el otro o viceversa. Por ejemplo el índice de simpatía y número enemigos que se tienen, o los tiempos de estudio y número de fracasos.



En la Estadística se suelen emplear muchos índices de correlación. Los dos más usuales son el cuantitativo de Pearson y el cualitativo de Spearman.



Charles Spearman

Karl Pearson

La cuantitativa, o de Pearson, se usa preferentemente con calificaciones numéricas y en ella se señalan valores o puntos a cada sujeto. Se refleja y expresa con el índice encerrado en la fórmula siguiente, que se recoge con el signo r (r latina) llamado coeficiente de Pearson.

$$r = \frac{n \cdot (\sum x \cdot y) - (\sum x) \cdot (\sum y)}{\sqrt{\left[n \cdot (\sum x^2) - (\sum x)^2 \right] \cdot \left[n \cdot (\sum y^2) - (\sum y)^2 \right]}}$$



Esto significa que $r = \frac{S \text{ de } x' \cdot S y'}{N \cdot Sgx' \cdot Sgy'}$

Entendiendo por [S de x'], la Suma de diferencias en el factor "x" respecto a la M.

Entendiendo por [S de y'], la Suma de las diferencias en el factor "y" respecto a la M.

Entendiendo por N, el numero de sujetos del grupo.

Entendiendo por Sgx', la Desviación típica del factor x.

Entendiendo por Sgy', la Desviación típica del factor y.

La correlación ordinal, o de Spearman, se expresa así con un p (r griega, r_o) y se usa preferentemente cuando se trabaja con cualidades y se señalan puestos en las series de los sujetos:

$$r_{rho} = \frac{1 - 3 \left(\frac{\sum D^2}{n(n^2-1)} \right)}{1 - 6 \left(\frac{\sum D^2}{n(n^2-1)} \right)}$$

O también así:

$$r_{s(xy)} = 1 - \frac{6 \cdot \sum d_1^2}{n \cdot (n^2-1)}$$

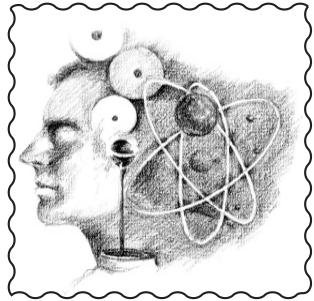


La fórmula significa;

$\rho = 1 - \frac{6 S d^2}{N \cdot (N^2 - 1)}$ $\frac{6 \text{ veces las diferencias al cuadrado}}{N^{\circ} \text{ sujetos por } N^{\circ} \text{ al cuadrado menos } 1}$

Significa el numerador: 6 veces la suma de las diferencias al cuadrado de puestos entre los factores "x" e "y" (por lo que se hace necesario determinar su puesto o colocación en la serie)

Significa el denominador el producto del número de miembros multiplicado por el mismo número al cuadrado menos 1.



Un ejemplo, para entender estos datos con facilidad, es el siguiente simplificado y sencillísimo, que puede ser ampliado por cualquiera en un trabajo casero. Resume más amplia explicación de cualquier libro elemental de Estadística, tanto descriptiva como inferencial (ver conceptos antes)

Sujeto	a	b	x'	y'	x2	y2	xy	P°a	P°b	d	d2
1	8	9	2	4	4	16	8	8°	10°	2	4
2	7	6	1	1	1	1	1	6°	5°	1	1
3	9	8	3	3	9	9	9	10°	9°	1	1
4	6	6	0	1	0	1	0	4°	5°	1	1
5	4	2	-2	-3	4	9	6	3°	2°	1	1
6	8	6	2	1	4	1	2	8°	5°	3	9
7	2	1	-4	-4	16	16	12	1°	1°	0	0
8	7	3	1	-2	1	4	-2	6°	4°	2	4
9	6	7	0	2	0	4	0	4°	8	4	16
10	3	2	-3	-3	9	9	9	2°	2°	0	0



Explicación de la tabla anterior:

- a. Factor a, por ejemplo tiempo de estudio valorado sobre 10.
- b. Factor b, por ejemplo resultados académicos valorados sobre 10.
- x' Diferencia en a respecto a la media que es 6.
- y' Diferencia en b respecto a la media que es 5.
- x^2 Diferencias en a, elevadas al cuadrado.
- y^2 Diferencias en b, elevadas al cuadrado.
- $x.y$ Producto de x por y.
- $P^{\circ}a$ Puesto que ocupa cada sujeto en a comenzando por el menor valor.
- $P^{\circ}b$ Puesto de cada sujeto en b comenzando por el menor valor.
- d Diferencia en puestos en a y en b.
- d^2 Diferencia de puestos elevada al cuadrado.

Se halla primero la Sigma o Desviación típica de cada serie, conforme al procedimiento indicado anteriormente. La fórmula sencilla era

Sd^2
O más vulgarmente dicho: $Sg : \sqrt{\frac{\quad}{N}}$



Es decir, en a:

Raíz cuadrada ($\sqrt{\quad}$) de $SD^2 (=47)$, dividido por N (10): Raíz cuadrada de $4,7 = 2,19$

Y en b:

Raíz cuadrada ($\sqrt{\quad}$) de $SD^2 (=70)$, dividido por N (10) :
Raíz cuadrada de $7 = 2,64$

Correlación de Pearson: r, en este ejercicio:

$$r = \frac{S \text{ de } x' \cdot S y'}{N \cdot Sg x' \cdot Sg y'} \text{ significa } \frac{x \cdot y : 49}{10 \times 2,19 \times 2,64 : 57,81} = + 0,84$$

$$P(ro) = p = 1 - \frac{6 S d^2}{N \cdot (N^2 - 1)} = 1 - \frac{6 \times 37}{10 \times (100 - 1)} = \frac{222}{990} = 1 - 0,22 = + 0,78$$

Con ambos coeficientes, (0,84 y 0,78) la correlación nos da un significado elevado entre los dos factores del ejemplo puesto como modelo. Se puede decir que entre Factor a, (tiempo de estudio valorado sobre 10) y el Factor b, (resultados académicos valorados sobre 10) hay una fuerte tendencia (correlación) a varias concomitantemente.

4. Las desviaciones típicas individuales

Son otro elemento estadístico que se suele usar con frecuencia. En referencia a las valoraciones individuales más técnicas, se emplea también en trabajos de investigación y precisión las llamadas desviaciones típicas individuales, Si, que son la relación entre la puntuación individual en referencia a la media de la serie o grupo, dividida esa diferencia por la desviación típica de la serie, Sg.

Se puede así situar a cada sujeto en una curva normalizada, o campana de Gauss, que hace referencia a la puntuación sigmática y a la frecuencia de sujetos que obtienen cada valor.

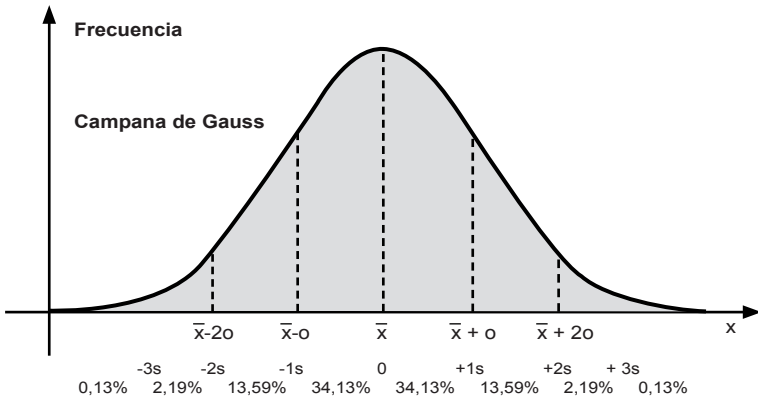
La formula de la desviación típica individual es:

Six = $\frac{X - M}{Sg}$ = $\frac{X - M}{Sg}$, valor del individuo y M, media de la serie
Sg = Desviación típica general de la serie

Six: Es la desviación típica del sujeto x.

X: Es el Valor del Sujeto X y M es la media de la serie.

Sg: Es la desviación típica general de la serie



Los estadísticos han realizado sus cálculos generales y lo normal es que se produzca un porcentaje típico, o estadísticamente idealizado, que responde a la curva que antecede, tanto más alto cuanto más los sujetos se acercan a la Media del grupo. A esta curva “normalizada” se denomina “campana de Gauss”, por el investigador matemático que la diseñó.



Carlos Federico Gauss

Ello significa que podemos considerar en cualquier rasgo, y a efectos estadísticos, que

- el 0,13 %** pertenece a la categoría muy superior (hacia + 4 Si)
- el 2,19 %** se muy sobresaliente (+ 3 Si)
- el 13,59 %** es bueno (+ 2 Si)
- el 34,13 %** es normal por encima de la media (0 a + 1 Si)
- el 34,13 %** es normal por debajo de la media (0 a - 1 Si)
- el 13,59 %** es malo (- 2 Si)
- el 2,19 %** es muy malo (- 3 Si)
- el 013 %** es muy deficiente o de categoría muy baja (hacia - 4 Si)

Esta escala significa que el 68,26%, los 2/3 de una población, deberán pertenecer estadísticamente a la normalidad. Y el 1/3 restante son más bien especiales o no normales, la mitad de ese 1/3 (1/6) por encima y la otra mitad (1/6) por debajo.

Quando se hacen trabajos de investigación, hay que saber moverse en estos lenguajes numéricos y matemáticamente comparativos, para que se puedan establecer referencias a las situaciones interdisciplinares e intercientíficas. Y no es conveniente

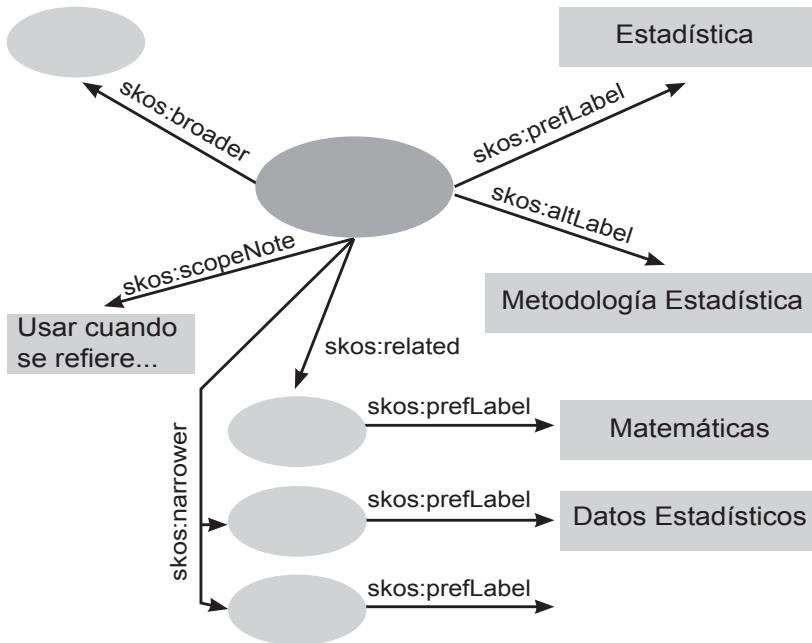


que todo lenguaje científico se asocie necesariamente con los procedimientos de laboratorio. Hay multitud de investigaciones que se pueden y deben realizar en los ambientes naturales de la vida cotidiana, con los cuales se puede objetivar muchas impresiones que en sí resultan parciales o muchas sospechas que reclaman aclaraciones.

5. Uso de los lenguajes científicos

Cuando un investigador redacta los resultados de sus trabajos, utiliza el lenguaje científico. Hay personas que confunden lenguaje científico con lenguaje oscuro. No es aceptable tal confusión. Es cierto que hay un lenguaje de expertos y otro de divulgación. Pero, en ambos, hay que tener en cuenta la claridad, la fundamentación y la proyección práctica.

Los conceptos estadísticos que se han puesto en las páginas anteriores no son propios de expertos, sino de gente medianamente culta. Cualquiera persona normal puede entenderlos, interpretarlos y usarlos.



Igualmente, cuando los alumnos de cualquier nivel estudian una determinada asignatura o realizan un trabajo escolar de investigación en cualquier disciplina, deben manejar con naturalidad el lenguaje estadístico. En una revista de divulgación se utiliza también este lenguaje, aunque matizado en sus elementos más crípticos por condicionantes periodísticos. En un prospecto de medicinas, en un manual de instrucciones, en un libro de texto, en una revista especializada, en la lista de ingredientes de una comida enlatada, en el informe de un radiólogo..., en todos estos textos, y en muchos más, se utiliza alguna variedad de lenguaje técnico o científico.

En realidad, un individuo de cultura media está constantemente en contacto con esta variedad; porque el lenguaje estadístico y técnico-científico no es más que una variedad del habla, que resulta de adaptar la lengua común a la comunicación de contenidos técnicos y numéricos.

Este lenguaje no es uniforme. Cada rama del saber, cada disciplina, utiliza un lenguaje propio. En las investigaciones educativas no se precisan muchos tecnicismos, pero sí los suficientes para que todos entiendan con claridad y concreción lo que se pretende exponer.

