

CONCEPCIONES INICIALES SOBRE LA ASOCIACIÓN ESTADÍSTICA

ESTEPA CASTRO, A.¹ y BATANERO BERNABEU, M.C.²

¹ Universidad de Jaén.

² Universidad de Granada.

SUMMARY

In this paper we make a summary of psychological investigation about statistics association. The results of experimental study about previous conceptions are shown. This study has comprised the students' judgments of association. Using factorial analysis, we have identified five factors, which determine the structure of students' strategies. We also show a qualitative study about students' strategies in 2×2 and $r \times c$ contingency tables, scatterplots, and comparison of a numerical variable in different samples. Finally we describe four different misconceptions about statistical association and the conclusions of this paper.

INTRODUCCIÓN

Entre las nociones estadísticas fundamentales, cuya enseñanza debe optimizarse, se encuentra la de asociación, que extiende la idea de correlación a variables cualesquiera, incluso no numéricas (Hildebrand et al. 1977). Una comprensión correcta de la misma es un prerrequisito básico para el aprendizaje de un numeroso conjunto de técnicas y métodos estadísticos. Es patente también la conexión entre los conceptos de asociación y causalidad, por lo cual, en el estudio de la asociación se apoya el método científico experimental y la toma de decisiones en situaciones de incertidumbre. Sin embargo, asociación y causalidad no siempre coinciden y, además, la detección de la asociación no es una habilidad espontáneamente adquirida. Los sujetos, aún adultos, se guían preferentemente por sus teorías previas sobre las interrelaciones entre variables, más que por los datos objetivos sobre las contingencias empíricas, con el consiguiente perjuicio en diversas áreas de la actividad humana.

En este trabajo presentamos los resultados de un estudio experimental sobre las concepciones previas de los alumnos y las estrategias empleadas en la resolución de problemas descriptivos de asociación estadística. El concepto de asociación es tratado en nuestro trabajo desde el punto de vista de la estadística descriptiva, sin plantearnos problemas de inferencia a otros datos diferentes de los que se presentan en cada una de las tareas propuestas.

Es parte de un estudio más amplio sobre el efecto de un experimento de enseñanza basado en la resolución de problemas y el uso del ordenador sobre el aprendizaje de este concepto (Estepa 1994) y esperamos que pueda contribuir a una mejor planificación de la enseñanza del tema.

INVESTIGACIONES PSICOLÓGICAS SOBRE ASOCIACIÓN ESTADÍSTICA EN TABLAS DE CONTINGENCIA 2×2

A pesar de la importancia en estadística del concepto de asociación y de las dificultades epistemológicas y psicológicas ligadas a este concepto, son escasas las investigaciones sobre el tema en didáctica de la matemática. La mayor parte de los estudios sobre asociación estadística se ha efectuado por psicólogos, que han analizado los juicios de asociación en tablas de contingencia 2×2 . Una tabla de contingencia se usa para presentar, en forma resumida, las frecuencias de casos en una población o muestra, clasificados respecto a dos variables estadísticas. El caso más simple es la tabla de contingencia 2×2 , en la cual las variables presentan sólo dos modalidades, con el formato que presentamos en la tabla I.

Tabla I
Esquema de la tabla de contingencia 2 x 2.

| | | | |
|-------|-------|-------|---------------|
| | B | no B | Total |
| A | a | b | a + b |
| no A | c | d | c + d |
| Total | a + c | b + d | a + b + c + d |

El estudio del razonamiento sobre la asociación estadística fue iniciado por Inhelder y Piaget (1955), quienes consideraron la comprensión de la idea de asociación como el último paso del desarrollo del concepto de probabilidad. Para ellos, los antecedentes evolutivos de los dos conceptos son los mismos, pues la comprensión de la idea de asociación implica las de proporción y probabilidad, y requiere capacidad combinatoria. Por este motivo sólo estudiaron el razonamiento sobre la asociación con sujetos de los estadios IIIa y IIIb correspondientes a las operaciones formales, proponiéndoles el problema de la asociación entre el color de ojos y del pelo, usando un conjunto de tarjetas coloreadas con dibujos de rostros como dispositivo experimental (cabello rubio y moreno, ojos azules y negros).

Con objeto de explicar sus resultados, Inhelder y Piaget clasificaron las cuatro posibilidades de combinar el color de pelo y ojos según el esquema presentado en la tabla I, en el cual *a*, *b*, *c* y *d* representan las frecuencias absolutas de las cuatro celdas de la tabla (cabello rubio, ojos azules; cabello rubio, ojos negros; cabello moreno, ojos azules y cabello moreno, ojos negros).

Según estos autores, para probar una asociación entre las dos variables, el sujeto deberá tener en cuenta los casos favorables (*a+d*), los casos desfavorables (*b+c*) y los casos posibles (*a+d*) + (*b+c*). Sin embargo, encontraron que, en el estadio IIIa, los adolescentes sólo analizan la relación del caso *a* (positivo-positivo) al total de datos, sin tener en cuenta el caso *d* (negativo-negativo), que equivale normativamente al anterior. Otras veces relacionarán *a* con *b* o con *c*, sin percatarse de que los casos de la celda *d* tienen el mismo valor normativo que los de la celda *a*. Enuncian así el problema de la asociación: *Superadas estas dificultades iniciales, el problema reside en comprender que la correlación no consiste en una simple probabilidad, es decir, una relación elemental entre los casos favorables (a+d) y el conjunto de los casos posibles (a+d+b+c). Sin duda alguna el problema verdadero de la correlación reside en este punto, y ello explica por qué la correlación sólo se adquiere en el nivel IIIb, aunque se posea la probabilidad simple en el nivel IIIa en su forma multiplicativa* (Inhelder y Piaget 1955, p. 198).

Estos autores consideran que el sujeto debe emplear la siguiente expresión para evaluar la asociación entre las variables: $R = [(a+d)-(b+c)] / [(a+d)+(b+c)]$. Opinan también que esta relación sólo se encuentra en el estadio IIIb, aproximadamente a los 15 años de edad.

Después de Inhelder y Piaget, numerosos psicólogos han estudiado los juicios de asociación, generalmente en

tablas de contingencia 2 x 2, con sujetos adultos, utilizando diferentes tipos de tareas y, como consecuencia, se ha hecho notar la pobre capacidad para establecer un juicio de asociación correcto. Por ejemplo, Smedlund (1863) y Shaklee y Mins (1982) encontraron sujetos adultos que basaban su juicio sólo en la celda *a*, o comparando *a* con *b*. La dificultad de este tipo de tarea se pone de manifiesto por el hecho de que, como hicieron notar Jenkins y Ward (1965), incluso la estrategia de comparar las diagonales de la tabla de contingencia, considerada correcta por Piaget e Inhelder, sólo es válida para tablas con frecuencias marginales iguales para la variable independiente.

Sin embargo, esta estrategia es ampliamente usada por adultos en las investigaciones de Allan y Jenkins (1983) y Shaklee y Tucker (1980). Para el caso general, Jenkins y Ward han propuesto como estrategia correcta la comparación de la diferencia entre las dos probabilidades $P(B|A)$ y $P(B|\bar{A})$, denominada $\delta P = [a/(a+b)] - [c/(c+d)]$.

Pérez Echeverría (1990) resume las estrategias que han sido identificadas en la investigación psicológica y las agrupa en cinco niveles de complejidad, calculando los porcentajes de los mismos en su investigación:

Nivel 1. Usar sólo una celda, usualmente la celda [a] (3,7% de uso).

Nivel 2. Comparar [a] con [b] o [a] con [c] (30,56%).

Nivel 3. Comparar [a] con [b] y [a] con [c] (15,74% de uso).

Nivel 4. Usar las cuatro celdas en la tabla, pero efectuando comparaciones aditivas (37,96% de uso).

Nivel 5. Deducir la asociación a partir de una relación multiplicativa entre las cuatro celdas de la tabla (12,04% de uso).

Otro aspecto de interés en este tema es que Chapman y Chapman (1969) mostraron la existencia de expectativas comunes acerca de las relaciones entre las variables que causan la impresión de contingencias empíricas. Este fenómeno ha sido llamado «correlación ilusoria», porque los sujetos mantienen sus creencias a pesar de la evidencia de la independencia de las variables. La experiencia vital y el entorno cultural de los sujetos contribuyen a la formación de estas teorías que se usan para interpretar los hechos y datos que les rodean. Jennings, Amabile y Ross (1982), Wright y Murphy (1984) y Alloy y Tabacnick (1984), entre otros, han estudiado este efecto de las teorías previas sobre el contexto del problema en los juicios de asociación.

La conclusión general es que, cuando hay acuerdo entre los datos y las teorías previas, la gente incrementa su confianza en la atribución de la covariación, pero cuando los datos no coinciden con estas teorías se produce un conflicto cognitivo y la precisión de la percepción de la covariación dependerá de la fuerza relativa de las dos fuentes de información.

METODOLOGÍA DEL ESTUDIO EXPERIMENTAL

Objetivos del estudio

Las investigaciones que hemos mencionado proporcionan una información necesaria y valiosa para la enseñanza del tema. Sin embargo, como hemos señalado, han estado centradas en tablas de contingencia 2 x 2. El concepto general de asociación lleva, sin embargo, a una clase más amplia de problemas, que intentamos cubrir en nuestra investigación. De este modo, podremos obtener una primera aproximación al estudio de las concepciones previas de los alumnos sobre la asociación estadística. Ello requiere una clasificación, desde el punto de vista matemático, de las estrategias empleadas en estos problemas y el análisis de las interrelaciones entre los juicios de asociación en los diferentes tipos de problemas, con objeto de estudiar la estructura general de los juicios de asociación.

Muestra

La muestra ha sido tomada intencionalmente y ha estado formada por 213 alumnos del Curso de Orientación Universitaria de tres institutos de bachillerato de la provincia de Jaén, donde 124 (58,2%) pertenecían al COU dirigido a ciencias y 89 (41,8%) al COU orientado a letras. De ellos 113 (53,1%) eran varones y el resto mujeres. Sólo 34 alumnos (16,0%) habían estudiado estadística en cursos anteriores. Únicamente 3 alumnos habían estudiado los conceptos de dependencia funcional y aleatoria. Como conclusión, podemos considerar que el estudio que realizamos respecto a la asociación estadística es una evaluación de las concepciones iniciales de los alumnos participantes (Artigue 1990, Confrey 1990).

Cuestionario

El cuestionario utilizado consta de diez problemas en los que el alumno ha de dar un juicio sobre la asociación existente entre dos variables, distribuidos en la forma siguiente: cinco problemas presentan los datos en forma de tablas de contingencia (tres tablas 2 x 2, una tabla 2 x 3 y otra 3 x 3); en otros tres se utiliza un diagrama de dispersión y los dos últimos son problemas de comparación de una variable numérica en dos muestras, uno de ellos empleando muestras relacionadas y el otro muestras independientes.

El cuestionario se presenta como apéndice. Los ítems sobre tablas 2 x 2 fueron tomados de la investigación de Pérez Echeverría (1990) modificando el formato y las frecuencias de algunas celdas. Los ítems 6 y 7 son datos reales y el resto de los ítems fue construido para esta investigación. En la elaboración del cuestionario se han tenido en cuenta las siguientes variables de tarea (Kilpatrick 1978):

V.1. Tipo de dependencia estadística entre las variables: contemplándose los tres posibles casos de dependencia directa, inversa e independencia en las tablas 2 x 2 y en los diagramas de dispersión. En las tablas 2 x 3 y 3 x 3 y en la comparación de muestras el signo de la asociación no es aplicable.

V.2. Intensidad de la dependencia. Para medir esta variable hemos utilizado el coeficiente de correlación de Pearson para las variables continuas, para las tablas de contingencia 2 x 2 el coeficiente PHI y para las demás tablas utilizamos el coeficiente V de Cramer. Para los dos últimos ítems se ha realizado un contraste de diferencias mediante el test t.

Tabla II
Valores de las variables en los ítems del cuestionario.

| VA- RIA- BLES | TABLAS DE CONTINGENCIA | | | | | NUBE DE PUNTOS | | COMPARACIÓN DE MUESTRAS | | |
|---------------------|------------------------|---------------------|----------------------|---------------------|----------------------|------------------|-----------------------|-------------------------|------------------|--------------------|
| | Tablas 2 x 2 | | Tablas r x c | | | Ítem 6 | Ítem 7 | Ítem 8 | Ítem 9 | Ítem 10 |
| | Ítem 1 | Ítem 2 | Ítem 3 | Ítem 4 | Ítem 5 | | | | | |
| V1 | Independencia | Dependencia inversa | Dependencia directa | Dependencia directa | Independencia | Independencia | Dependencia inversa | Dependencia directa | p = 0,0092 | p = 0,1510 |
| V2 | 0 | -0,44 | 0,67 | 0,37 | 0,1 | 0,11 | -0,77 | 0,55 | t = 3,3 | t = 1,5 |
| V3 | Sí | Sí | No | Sí | No | Sí | Sí | No | | |
| V4 | No | Sí | No hay teoría previa | Sí | No hay teoría previa | No | Sí | No hay teoría previa | Sí | No |
| V5 | Causal unilateral | Causal unilateral | Covariación casual | Causal unilateral | Interdependencia | Interdependencia | Dependencia indirecta | Concordancia | Interdependencia | Covariación casual |

V.3. Diferenciación de la variable dependiente e independiente. En ciertos problemas de asociación se puede distinguir claramente la variable dependiente de la variable independiente, mientras que en otros esta distinción no es posible.

V.4. Concordancia entre los datos y las teorías previas sugerida por el contexto del problema. En las tres categorías de preguntas se han propuesto ítems donde unas veces los datos concuerdan con las teorías previas y otras no.

V.5. Tipo de covariación. Se toman las cinco categorías que se contemplan en Barbancho (1973) *dependencia causal unilateral, interdependencia, dependencia indirecta, concordancia y covariación casual.*

En la tabla II se muestra un resumen de los valores de las variables de tarea utilizados en cada uno de los ítems. El cuestionario fue probado con una muestra piloto adicional de 51 estudiantes, realizándose un estudio de generalizabilidad (Brennan 1983). Se obtuvo un índice de generalizabilidad $G_1 = 0,84$ como medida de la posibilidad de extensión de nuestras conclusiones a la población de ítems hipotéticos y otro $G_2 = 0,80$ de generalizabilidad a la población de estudiantes. El resto del estudio se ha efectuado sobre la muestra de 213 alumnos que hemos descrito con anterioridad.

JUICIOS DE ASOCIACIÓN

Las frecuencias de juicios correctos e incorrectos, dados por los alumnos en cada uno de los ítems del cuestionario, se muestra en la tabla III.

En el ítem 2, correspondiente a la dependencia inversa en una tabla 2 x 2, un porcentaje apreciable de alumnos ha considerado que existe independencia entre las variables. Este tipo de respuestas fue también encontrado por Jenkins y Ward (1965) en tareas presentadas a los sujetos en un contexto causal. Para explicarla, observaron la falta de simetría entre las variables que se consideran como causas y las que actúan como efectos. Si existen más casos en que, dándose la causa, no se presenta el efecto, los sujetos concluirán que causa y efecto no se relacionan, esto es, creerán que existe independencia entre las variables. Interpretamos este tipo de respuesta como indicativa de que los estudiantes que la proporcionan tienen una *concepción unidireccional* de la asociación. Consideran sólo la asociación entre dos variables estadísticas cuando el signo de la misma es directo e interpretan la asociación inversa como independencia.

Esta concepción puede ser también identificada en el desarrollo histórico del tema. Como Pearson (1920) señaló, el mismo Galton no fue consciente de la posibilidad de la asociación inversa cuando presentó su definición de correlación ante la Royal Society el 5 de diciembre de 1888: *dos órganos variables se dicen que están correlacionados cuando la variación de uno de estos órganos es seguida por una variación del otro en la misma dirección* (p. 199). Pearson completa esta información exponiendo que el significado de un coeficiente de correlación negativo fue debido a Weldon, quien publicó en 1892 un artículo relacionado con este tema.

Otro hecho notable es la diferencia en dificultad en los tres primeros ítems, referidos a tablas de contingencia 2 x 2. En los dos primeros, el contexto puede sugerir al alumno una relación entre la variable diferente a la que se presenta en los datos, especialmente en el ítem 1, correspondiente a la independencia, en el que el contex-

Tabla III
Frecuencia y porcentaje de respuestas sobre el juicio de asociación.

| Ítem | Juicio correcto | Juicio incorrecto | No responde | Total |
|------|-----------------|-------------------|-------------|-----------|
| 1 | 84 (39,4) | 119 (55,7) | 10 (4,7) | 213 (100) |
| 2 | 108 (50,7) | 90 (42,2) | 15 (7,0) | 213 (100) |
| 3 | 194 (91,1) | 12 (5,6) | 7 (3,3) | 213 (100) |
| 4 | 186 (87,7) | 19 (8,9) | 8 (3,8) | 213 (100) |
| 5 | 129 (60,6) | 23 (10,8) | 61 (28,6) | 213 (100) |
| 6 | 177 (83,1) | 32 (15,0) | 4 (1,9) | 213 (100) |
| 7 | 182 (85,4) | 30 (14,1) | 1 (0,5) | 213 (100) |
| 8 | 46 (21,6) | 155 (72,8) | 12 (5,6) | 213 (100) |
| 9 | 178 (83,6) | 20 (9,4) | 15 (7,0) | 213 (100) |
| 10 | 157 (73,7) | 44 (20,7) | 12 (5,6) | 213 (100) |

to sugiere a los alumnos la dependencia de las variables (fumar y padecer trastornos bronquiales). Como señala Crocker (1981), detectar la interrelación entre los acontecimientos o valorar las contingencias ambientales es algo más que un problema matemático. Es una tarea en que se ejercita nuestro conocimiento previo del fenómeno y en el cual influye poderosamente el contexto de la información que estamos empleando.

En el ítem 1 es preciso además un empleo correcto del razonamiento proporcional. Este ítem puede verse también como un problema de comparación de dos probabilidades en la que el número de casos favorables y desfavorables en los términos de comparación es proporcional. Para Inhelder y Piaget (1955) corresponde al nivel IIIb de razonamiento probabilístico y es la tarea más difícil dentro de las de comparación de probabilidades. Todos estos aspectos se discuten con más detalle en Batanero y otros (en prensa).

Análisis factorial de respuestas en los juicios de asociación

Puesto que las respuestas de cada alumno particular sobre la existencia o no de asociación entre las variables dependerá de sus concepciones, cabe esperar una relación entre las respuestas de un mismo alumno. Para determinar la estructura de este conjunto de respuestas se ha efectuado un análisis factorial usando el paquete estadístico BMDP. La extracción inicial de factores se ha realizado con el método del factor principal y como método de rotación se ha empleado la rotación varimax (Cuadras 1991).

En la tabla IV se presentan los autovalores y comunalidades obtenidos. De la estructura de los autovalores, puede observarse la multidimensionalidad del juicio de asociación: ninguno de los factores destaca por el porcentaje de inercia respecto a los demás y no se logra reducir la dimensión del espacio vectorial subyacente.

Tabla IV
Autovalores y comunalidades del análisis factorial.

| Ítem variable | Comunalidad | Factor | Autovalor | Porcentaje de varianza | Porcentaje de varianza acumulado |
|---------------|-------------|--------|-----------|------------------------|----------------------------------|
| 1 - D6 | 0,52928 | 1 | 1,68828 | 16,9 | 16,9 |
| 2 - E2 | 0,65924 | 2 | 1,27656 | 12,8 | 29,6 |
| 3 - F2 | 0,65310 | 3 | 1,15154 | 11,5 | 41,2 |
| 4 - G2 | 0,65948 | 4 | 1,06427 | 10,6 | 51,8 |
| 5 - H2 | 0,46009 | 5 | 1,02437 | 10,2 | 62,1 |
| 6 - I3 | 0,54413 | | | | |
| 7 - J1 | 0,53643 | | | | |
| 8 - K1 | 0,76092 | | | | |
| 9 - L1 | 0,72373 | | | | |
| 10 - M1 | 0,67863 | | | | |

Tabla V
Matriz factorial rotada por el método varimax.

| Ítem | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|------|----------|----------|----------|----------|----------|
| 1 | 0,69014 | | | | |
| 2 | | | | | 0,80831 |
| 3 | 0,56512 | | -0,42876 | | 0,36397 |
| 4 | 0,50840 | | | -0,55696 | |
| 5 | 0,53138 | | 0,37402 | | |
| 6 | | | 0,69078 | | |
| 7 | | 0,36823 | 0,54935 | | |
| 8 | | | | 0,83021 | |
| 9 | | 0,80612 | | | |
| 10 | | 0,68755 | | | -0,37770 |

Hay cinco factores con inercia superior a la unidad y la importancia relativa de los mismos es muy parecida. El conjunto de los cinco sólo explica un 62% de la varianza y cada ítem presenta un porcentaje apreciable de varianza no explicada por el conjunto de factores, lo que indica su fuerte especificidad. Este hecho, a nuestro juicio, confirma los resultados de los diferentes trabajos psicológicos sobre juicios de asociación en tablas de contingencia y los extiende al caso de tareas más complejas. El juicio de asociación no viene determinado por una variable de tarea única; las diferentes variables interaccionan, provocando en el sujeto tanto un cambio de estrategia como un cambio en el juicio proporcionado.

En la tabla V presentamos la matriz factorial obtenida tras la rotación varimax, en la que hemos suprimido, para una mejor interpretación, los coeficientes factoriales de magnitud pequeña. A continuación interpretamos los factores obtenidos.

Factor 1: Asociación entre variables como variación de las frecuencias relativas condicionales.

Agrupamos todas las tablas de contingencia, excepto la tabla del ítem 2. Por ello, el primer factor expresa la diferencia del tipo de habilidad requerida en los juicios de asociación en tablas, respecto a los otros ítems, debido al hecho de tener que trabajar con las frecuencias de la tabla, ya que no es posible emplear valores numéricos para intentar relacionar las variables. En estos ítems es preciso concebir la asociación estadística como la variación de las frecuencias relativas condicionales de una de las variables en función de la otra.

Factor 2: Asociación entre variables como variación del valor numérico de una de ellas en función de la otra variable.

Agrupamos el ítem 7 relativo a la dependencia inversa en una nube de puntos en el cual la intensidad de la asociación y la pendiente acusada de la nube de puntos permite apreciar visualmente el decrecimiento de la gráfica y los dos ítems referidos a la comparación de una variable numérica en dos muestras (9 y 10). En estos tres ítems (y sólo en ellos) se observa claramente cómo el valor numérico de la variable dependiente varía en función del valor de la variable independiente.

Factor 3: Signo de la asociación cuando no hay teorías en contra.

Opone el ítem 3, que tiene una dependencia directa fuerte, frente a los ítems 5, 6 (independencia) y 7 (dependencia inversa fuerte). En ninguno de los casos el contexto de los ítems sugiere una asociación diferente a la presentada en los datos. En el resto de los ítems o bien la teoría está en contra de los datos o no tiene sentido considerar el signo de la asociación. Concluimos en consecuencia que este factor representaría el signo de la asociación, que será tenido en cuenta al producir un juicio, diferenciando la asociación de tipo directo frente a la independencia o dependencia inversa, en el caso de

que las teorías previas no contradigan los datos y haya una intensidad fuerte en la asociación.

Factor 4: Concordancia frente a causalidad (correlación moderada)

Aparecen con signos opuestos dos ítems en que la intensidad de la asociación es moderada, diferenciándose por el tipo de relación entre las variables a la que es atribuible la asociación. En el ítem 4 (número de horas y éxito en un examen), el alumno puede atribuir a la variable independiente un papel de causa y a la dependiente el de efecto, pues su experiencia le dice que, en general, un mayor número de horas de estudio aumenta la probabilidad de éxito en un examen. Aunque la correlación sea moderada, al darse la coincidencia entre correlación y causalidad en este contexto, el alumno produce un juicio positivo de asociación.

Por el contrario, en el ítem 8 referido a la concordancia en las puntuaciones dadas a una serie de concursantes por dos jueces, no hay una relación de causa-efecto. Aparece una dificultad semántica derivada del hecho de que las palabras que empleamos en el lenguaje matemático pueden tener un sentido distinto en la vida cotidiana. Aunque la pregunta se hace evitando la palabra dependencia en este caso, puede ocurrir que el alumno asocie el término «relación» empleado en la pregunta con la idea de dependencia y ello provoque la confusión mostrada.

Factor 5: Asociación como comparación de muestras independientes.

En este factor aparecen agrupados, aunque con signo opuesto, las dos tablas de contingencia 2 x 2 y el último ítem que se refiere a la comparación de muestras independientes. En los tres casos se trata de comparar dos muestras independientes y existe asociación. De este modo, interpretamos este factor como correspondiente a una concepción de la asociación entre una variable cualquiera y otra dicotómica como la existencia de diferencias en las frecuencias o valores de la variable en ambas muestras.

ESTUDIO CUALITATIVO DE ESTRATEGIAS

Para completar nuestro estudio, se han clasificado las estrategias de los 213 estudiantes de la muestra en cada uno de los problemas de asociación atendiendo a su corrección desde el punto de vista de los conceptos matemáticos puestos en juego implícita o explícitamente. Para ello las respuestas escritas de los alumnos fueron sometidas a un proceso de análisis de contenido y categorizadas en función de la información utilizada y el modo en que esta información se combina o compara tras una serie de procesos de comparación de las respuestas similares. Finalmente a partir de las categorías iniciales, la comparación con las estrategias descritas en las investigaciones previas y el asesoramiento de otros investigadores que colaboraron en este proceso se llegó a la

definición de las estrategias que presentamos en este trabajo.

Vergnaud (1990), al discutir la actividad realizada por un alumno durante el proceso de resolución de problemas, distingue dos tipos de situaciones: a) situaciones para las que el sujeto dispone en su repertorio, en un momento dado, de las competencias necesarias para un tratamiento relativamente rápido de la misma; b) situaciones para las cuales el sujeto no dispone de dichas competencias.

Este autor llama *esquema* a la organización invariante de la conducta para una clase de situaciones dadas: lo considera útil en los dos tipos de situaciones descritas. En las primeras el sujeto tendrá una conducta bastante automatizada y tenderá a usar un esquema único. En el segundo tipo de situaciones se observará el empleo sucesivo de varios esquemas que pueden entrar en competición y que deben ser combinados y descombinados para llegar a la solución buscada; si ésta es correcta, se ha producido un aprendizaje. En los esquemas, Vergnaud sitúa los conocimientos en acto del sujeto (conceptos y teoremas en acto). El análisis de estos *esquemas* latentes tras las estrategias descubre elementos cognitivos implícitos, esto es, conceptos y teoremas en acto, según este autor.

Para resolver los problemas propuestos, los estudiantes de nuestra muestra estarían ante el segundo tipo de situaciones descrito por Vergnaud. En las estrategias identificadas en nuestro estudio se han observado coincidencias de grupos importantes de alumnos en su proceso de resolución de un mismo tipo de problemas. En consecuencia, creemos adecuado aplicar el marco conceptual propuesto por Vergnaud para el análisis de estas estrategias. A continuación describimos las estrategias empleadas en los distintos tipos de problemas propuestos y los teoremas en acto identificables en las mismas. Estos teoremas en acto no son propiamente teoremas cuyo enunciado se ajuste exactamente a los que aparecen en los libros de texto, sino propiedades y relaciones de los objetos matemáticos usados implícitamente.

Estrategias empleadas en tablas de contingencia

Como hemos indicado, la tabla de contingencia ha sido el formato más empleado en las investigaciones sobre juicios de asociación en el campo de la psicología. Aunque en Estepa (1994) se han comparado los resultados obtenidos con los de otras investigaciones, en este trabajo preferimos presentar una clasificación original, que tiene en cuenta los elementos conceptuales matemáticos utilizados implícitamente por los alumnos.

Una característica importante de la tabla de contingencia es que el alumno debe trabajar con las frecuencias contenidas en la misma para obtener su juicio de asociación. Al tratarse de variables cualitativas no pueden establecer relaciones numéricas entre las mis-

mas. La clasificación de las estrategias, que ha partido del análisis del modo en que el alumno emplea las distintas frecuencias de la tabla, se presenta a continuación.

Estrategias correctas

ST.1. *Comparación de todas las distribuciones de frecuencias relativas condicionales $h(B_j|A_i)$ de cada valor B_j en dos o más valores diferentes de A_i .* En este caso el alumno emplearía implícitamente el siguiente teorema en acto T1: «La dependencia de una variable B respecto a otra variable A implica la variación de las frecuencias relativas condicionales $h(B_j|A_i)$ cuando A_i varía». Una interpretación similar tendría el caso en que el alumno intercambiase filas por columnas.

ST.2. *Comparación de una sola distribución de frecuencias relativas condicionales $h(B_j|A_i)$ para un valor fijo j , en cada posible valor de A_i , con la frecuencia relativa marginal h_j .* Estos alumnos usarían implícitamente el teorema en acto T2: «La independencia implica la invarianza de la distribución de B al condicionarla con un valor de A». (Podemos intercambiar filas por columnas.)

ST.3. *Comparación de posibilidades.* Cuando el estudiante, en lugar de calcular las probabilidades, compara las frecuencias de casos a favor y en contra de B (o la razón de estas frecuencias) en cada valor de A. Se emplea implícitamente el teorema T3; «hay una correspondencia unívoca entre la probabilidad de un suceso y la razón de sus posibilidades a favor y en contra».

Estrategias parcialmente correctas

ST.4. *Comparación de dos o más distribuciones de frecuencias absolutas condicionales con la frecuencia absoluta marginal correspondiente.* Es similar a la estrategia ST.1, pero no se emplean explícitamente las frecuencias relativas. En lugar de ello se hacen comparaciones aditivas o cualitativas, con lo que no se cuantifican correctamente las probabilidades implicadas.

ST.5. *Comparación de una única frecuencia absoluta en cada distribución condicional con la correspondiente frecuencia marginal.* Es similar a ST.4, aunque se emplea sólo una frecuencia absoluta en cada distribución condicional. Se emplea implícitamente el teorema en acto T4; «es suficiente la variación de la frecuencia relativa de un sólo valor de A cuando B varía, para probar la dependencia».

ST.6. *Comparación de la suma de frecuencias en las diagonales.* Es la estrategia descrita por Inhelder y Piaget, válida sólo cuando la tabla tiene iguales frecuencias marginales para la variable independiente. Algunos estudiantes emplean también esta estrategia en tablas 2×3 o 3×3 , reduciéndola previamente a una tabla 2×2 sumando algunas filas o suprimiéndolas.

Estrategias incorrectas

ST.7. Usar sólo una celda de la tabla, en general, la de máxima frecuencia. Hemos relacionado esta estrategia con una concepción que denominaremos **localista** de la asociación. El estudiante no usa toda la información relevante al problema, sino sólo la más sobresaliente. También puede estar relacionada con la heurística de la representatividad descrita por Kahneman y otros (1982). En el caso de la tabla 2x2, a veces los estudiantes argumentan que no existe dependencia entre las variables porque una de las celdas [b] o [c] no tiene frecuencia nula. Ambos casos indican también una **concepción determinista de la asociación**, porque exigen que, siempre que se produzca A (causa), se produzca B (efecto), o bien que no pueda producirse B a menos que se produzca A.

ST.8. Usar sólo una distribución condicional. Estos estudiantes no conciben el problema de asociación como un problema de comparación de probabilidades, basando su juicio sólo en un valor fijo de la variable A. Por tanto, también indicaría la concepción localista de la asociación.

ST.9. Comparar algunas frecuencias dobles con el número total de observaciones, en particular las de máxima frecuencia. Similar al caso ST.7.

ST.10. Comparar frecuencias marginales; algunos estudiantes consideran que no se puede llegar a una conclusión si las frecuencias marginales de la tabla son diferentes.

ST.11. Otros procedimientos. Hemos catalogado aquí otros procedimientos que aparecen esporádicamente y que no se relacionan con las frecuencias de la tabla.

Finalmente, en la tabla VI presentamos las frecuencias y porcentajes con que se han usado las estrategias descritas en las tablas de contingencia. Puede observarse la diferenciación de estrategias al pasar de la tabla 2 x 2 a las de dimensión superior (ítems 4 y 5). Asimismo es notable en el ítem 5 la falta de respuesta. Deducimos de ello que, aunque un porcentaje de alumnos apreciable es capaz intuitivamente de usar estrategias correctas o parcialmente correctas en tablas 2 x 2 no siempre son capaces de extenderlas a los casos más complejos.

Estrategias empleadas en diagramas de dispersión

En un diagrama de dispersión, la información sobre la covariación de las variables se presenta al alumno en forma de diagrama cartesiano de un conjunto de pares de valores (x_i, y_i) . No es natural el trabajo directo con las distribuciones condicionales de frecuencias relativas. Por el contrario, a partir de la representación gráfica puede determinarse visualmente la asociación entre las dos variables. A continuación comentamos las estrategias detectadas.

Estrategias correctas

SN.1. Comparación global. El alumno efectúa una comparación global correcta de la relación entre las dos variables. Generalmente se aplica implícitamente el teorema T2 anteriormente descrito.

SN.2. Crecimiento. Se usa como argumento el crecimiento, decrecimiento o constancia de la forma de la nube de puntos para justificar el tipo de dependencia. Se emplea implícitamente el teorema T5: «el crecimiento / decrecimiento de la nube de puntos indica la existencia de una correlación directa / inversa de las variables».

Tabla VI
Distribución de estrategias en tablas de contingencia.

| Estrategia | Ítem 1 | Ítem 2 | Ítem 3 | Ítem 4 | Ítem 5 |
|-------------|-----------|-----------|-----------|-----------|-----------|
| ST.1 | 8 (3,8) | 16 (7,5) | 2 (0,9) | 20 (9,4) | 18 (8,4) |
| ST.2 | 33 (15,5) | 27 (12,6) | 35 (16,4) | 21 (9,9) | |
| ST.3 | 13 (6,1) | 39 (18,3) | 34 (16,0) | 62 (29,1) | 5 (2,3) |
| ST.4 | 3 (1,4) | 19 (8,9) | 21 (8,9) | 15 (7,0) | 76 (35,6) |
| ST.5 | 2 (0,9) | 19 (8,9) | 17 (8,0) | | |
| ST.6 | 5 (2,3) | 11 (5,2) | 18 (8,4) | 7 (3,3) | 4 (1,9) |
| ST.7 | 29 (13,6) | 26 (12,2) | 22 (10,3) | 6 (2,8) | |
| ST.8 | 56 (26,3) | 28 (13,2) | 30 (14,1) | 47 (22,1) | 24 (11,3) |
| ST.9 | 1 (0,5) | 6 (2,8) | 7 (3,3) | 2 (0,9) | |
| ST.10 | 8 (3,8) | | 1 (0,5) | 6 (2,8) | |
| ST.11 | 13 (6,1) | 2 (0,9) | 3 (1,4) | 7 (3,3) | 4 (1,9) |
| No contesta | 42 (19,7) | 20 (9,8) | 23 (10,8) | 20 (9,4) | 82 (38,5) |
| Total | 213 (100) | 213 (100) | 213 (100) | 213 (100) | 213 (100) |

Estrategias parcialmente correctas

SN.3. Comparación con un patrón. El alumno compara la forma de la nube de punto con una función conocida, por ejemplo, lineal o cuadrática, para argumentar correctamente la relación entre las variables. Se emplea implícitamente el teorema T6: «si es posible ajustar una función no constante a una nube de puntos, en forma en que la aproximación obtenida sea aceptable, existirá una asociación moderada o alta entre las variables».

SN.4. Interpretación correcta de puntos aislados. Se citan pares concretos de puntos que cumplen el tipo de relación existente entre las variables. Aunque el alumno los interprete correctamente, es una estrategia incompleta por no emplear toda la información disponible.

Estrategias incorrectas

SN.5. Interpretación incorrecta de puntos aislados. El alumno emplea pares aislados de valores para interpretar, de forma incorrecta, la relación entre las variables.

SN.6. Teorías previas. Se usan las teorías previas sobre el contexto para argumentar la asociación, a pesar de la evidencia de los datos; es decir, se manifiesta explícitamente la *correlación ilusoria*.

SN.7. Otras variables. Cuando la existencia de otras variables que puedan influir en la dependiente es considerada como motivo para la no existencia de asociación. Estos alumnos interpretan la pregunta sobre la dependencia en términos de relación causa-efecto, manifestando una concepción causalista de la asociación; exigen además la eliminación de otras explicaciones causales del fenómeno para admitir la existencia de relación.

SN.8. Uniformidad. Se argumenta que no existe dependencia porque a un solo valor de la variable dependiente pueden corresponder varios de la dependiente, ya que la relación no es de tipo funcional. Consideramos que estos alumnos manifiestan la concepción determinista de la asociación. En particular algunos alumnos exigen la existencia de una expresión algebraica que ligue las variables.

SN.9. Causalidad. A pesar de la asociación empírica, se argumenta que no existe asociación entre las variables, ya que la variable, independiente no puede causar una influencia directa sobre la dependiente; es decir no existe relación causa-efecto. Esta estrategia se ha observado con relación al ítem 8 que presenta una correlación debida a la concordancia entre las puntuaciones dadas independientemente por dos jueces a una serie de concursantes a un premio. En consecuencia, estos alumnos manifiestan una *concepción causalista de la asociación*, identificando asociación y causalidad.

SN.10. Otras. Estrategias no contempladas anteriormente.

Finalmente, en la tabla VII presentamos la distribución de estrategias en los diagramas de dispersión. Observamos que el estudio del crecimiento de la función ha sido especialmente relevante en el ítem 7 en el que la correlación es intensa. Por el contrario, en el caso de independencia (ítem 6), se han empleado preferentemente estrategias incorrectas, como argumentar la presencia de otras variables que influyan sobre la dependiente, para justificar la falta de asociación entre las variables.

Estrategias empleadas en la comparación de muestras

En este caso, el alumno puede trabajar con los valores numéricos de la variable dependiente, aunque no posee

Tabla VII
Distribución de las estrategias en diagramas de dispersión.

| ESTRATEGIAS | Ítem 6 | Ítem 7 | Ítem 8 |
|--|-----------|------------|------------|
| Correctas | | | |
| SN.1. Comparación global | 16 (7,5) | 3 (1,4) | 28 (13,19) |
| SN.2. Crecimiento | 15 (7,0) | 134 (62,9) | 8 (3,8) |
| Parcialmente correctas | | | |
| SN.3. Compara con un patrón | | | 24 (11,3) |
| SN.4. Interpretación correcta de puntos aislados | 49 (23,0) | 27 (12,7) | 4 (1,9) |
| Estrategias incorrectas | | | |
| SN.5. Interpretación incorrecta de puntos aislados | 4 (1,9) | 3 (1,4) | 25 (11,7) |
| SN.6. Teorías previas | 22 (10,3) | 15 (7,0) | 6 (2,8) |
| SN.7. Otras variables | 68 (31,9) | 1 (0,5) | |
| SN.8. Uniformidad | 14 (6,6) | 1 (0,5) | 21 (9,9) |
| SN.9. Causalidad | | | 50 (23,5) |
| SN.10. Otras | 10 (4,7) | 8 (3,8) | 6 (2,8) |
| Argumento confuso. No contesta | 15 (7,0) | 21 (9,9) | 41 (19,3) |

una representación gráfica que le permita apreciar visualmente la asociación. También puede trabajar con las distribución de frecuencias condicionales, pero ello requerirá la construcción de una tabla de contingencia, ya que los valores en cada muestra se le presentan individualmente. Las estrategias halladas se describen a continuación.

Estrategias correctas

SM.1. Comparación de las medias. El alumno compara las medias de las dos muestras para decidir el tipo de relación entre las variables. Se emplea implícitamente el teorema en acto T7: «Si las medias de la distribución condicional de una variable cuantitativa respecto a otra cualitativa varían, es porque existe asociación entre las variables.»

SM.2. Comparación de los totales. Los alumnos suman los valores de la variable dependiente en cada muestra, comparando las sumas obtenidas para argumentar el tipo de dependencia. Se emplearía implícitamente una versión modificada del teorema T7, aunque no se llega a calcular las medias.

SM.3. Comparación porcentajes en casos en que se cumple o deja de cumplirse el tipo de relación esperada en cada una de las dos muestras. Estos alumnos comparan las posibilidades a favor y en contra de un cierto tipo de asociación, habiendo transformado el problema en el estudio de la asociación en una tabla de contingencia 2 x 2. Se estaría implícitamente empleando el teorema T1.

SM.4. Comparación de las distribuciones. Se comparan las frecuencias o porcentajes de casos para cada uno de los valores de la variable dependiente en las dos muestras. También en este caso se emplearía uno de los teoremas T1 o T2.

Estrategias parcialmente correctas

SM.5. Comparación caso a caso. Se comparan los valores emparejados de las variables en las muestras relacionadas. Consideramos esta estrategia parcialmente correcta, porque no llegan a utilizar los estadísticos o las distribuciones de la variable en las dos muestras. Se emplearía implícitamente el teorema en acto T8: «La dependencia de los valores de una misma variable en dos muestras relacionadas implica que la diferencia entre las mismas ha de tener una media no nula.»

SM.6. Casos excepcionales. Se analizan los valores de las dos variables, resaltando los casos en que no se cumple la relación. No se llega a calcular porcentajes o proporciones.

SM.7. Halla las diferencias de valores correspondientes en muestras relacionadas.

SM.8. Comparación global.

Estrategias incorrectas

SM.9. Valores similares de la variable dependiente en las dos muestras. Si no los encuentra, considera que no existe relación entre las dos variables. Estos alumnos manifiestan la concepción determinista de la asociación.

SM.10. Comparación de máximos en ambas distribuciones, basando su juicio sólo en ellos, lo que consideramos indicativo de la concepción localista de la asociación.

SM.11. Comparación de recorridos de las dos distribuciones, basando su juicio en la similitud de los mismos.

SM.12. Coincidencias de los valores de las variables como signo de independencia.

SM.13. Teorías previas sobre el contexto para decidir el tipo de asociación, en lugar de emplear los datos.

SM.14. Otras estrategias no incluidas en las anteriores.

En la tabla VIII presentamos la frecuencia de uso de estas estrategias en el ítem 9 (muestras relacionadas) y 10 (muestras independientes). Observamos que los alumnos no emplean indistintamente las mismas estrategias en ambos ítems, habiendo diferenciado estos tipos de muestras.

Tabla VIII
Estrategias empleadas en la comparación de muestras.

| ESTRATEGIAS | Ítem 9 | Ítem 10 |
|----------------------------------|-----------|-----------|
| Correctas | | |
| SM.1. Compara las medias | 2 (0,9) | 26 (12,2) |
| SM.2. Compara los totales. | 1 (0,5) | 35 (16,4) |
| SM.3. Compara porcentajes | 41 (19,3) | 1 (0,5) |
| SM.4. Compara las distribuciones | | 34 (16,0) |
| Parcialmente correctas | | |
| SM.5. Compara caso a caso | 73 (34,3) | 3 (1,4) |
| SM.6. Señala casos excepcionales | 63 (29,6) | |
| SM.7. Halla las diferencias | 6 (2,8) | |
| SM.8. Comparación global | | 64 (30,0) |
| Incorrectas | | |
| M.9. Espera valores similares | 3 (1,4) | 1 (0,5) |
| SM.10. Compara máximos | | 6 (2,8) |
| SM.11. Compara recorridos | | 3 (1,4) |
| SM.12. Valora coincidencias | 1 (0,5) | 3 (1,4) |
| SM.13. Expone teorías previas | 1 (0,5) | 8 (3,8) |
| SM.14. Otras | 8 (3,8) | 14 (6,6) |

CONCEPCIONES PREVIAS SOBRE LA ASOCIACIÓN ESTADÍSTICA

Como hemos indicado a lo largo de la exposición, en los juicios de asociación y estrategias utilizadas, hemos identificado una serie de concepciones que los alumnos

poseen antes de la instrucción y que hemos denominado: *causalista, determinista, unidireccional y localista*. A continuación se describen las características de estas concepciones.

Concepción causalista

Hay alumnos que sólo consideran la dependencia entre variables si puede atribuirse a la existencia de una relación de tipo causal entre las mismas. Esto se ha hecho notar en especial en el ítem 8, en el cual la asociación se debe a la concordancia entre las variables y no a la influencia directa, de tipo causal, de una sobre otra. Los alumnos que manifiestan esta concepción argumentan que no existe relación entre las variables debido a que el valor de una de ellas no ejerce una influencia sobre la otra: *Observando la gráfica yo pienso que no existe ninguna relación entre la puntuación dada por cada juez. Cada uno ha tenido distinto gusto, tal vez han coincidido en alguno de ellos pero no existe relación entre los dos al igual que tampoco hay contradicción; cada uno de ellos tiene distinta opinión.*

En otros casos, como el ítem 6, la existencia de otras variables que influyan sobre la que el alumno toma como dependiente es vista como falta de asociación entre las variables. Se manifiesta aquí tanto una concepción causalista extrema, que implica el desechar otras posibles causas que produzcan un mismo efecto, como la concepción determinista que describimos después. El sujeto emplearía incorrectamente el principio de discriminabilidad de Kelley (otras causas son seguidas por el efecto) olvidando el principio de la consistencia (la causa es seguida por el efecto) (Pozo 1987): *No hay relación porque la tabla debería hacerse también con los partidos ganados y perdidos de cada equipo. Un ejemplo, el número 4 de la clasificación empató 12 partidos, sin embargo, el número 20 con 5 empates está el último[...] Porque el puesto final en la liga no se consigue sólo con los puntos sumados de los empates, también se consigue con los puntos que gana al perder partidos o al ganarlos.*

Concepción determinista

Algunos alumnos no admiten excepciones a la relación existente entre las variables, esperando una correspondencia que a cada valor de la variable independiente asigne un sólo valor de la dependiente, esto es, una relación de tipo determinista o funcional. Esta concepción se manifiesta en la forma siguiente:

En las tablas de contingencia 2 x 2, algunos alumnos argumentan que una de las celdas *b* y *c*, o ambas, de la tabla debieran tener frecuencia cero. En caso de que esto no ocurra, consideran que ello es debido a que no existe asociación entre las variables, esto es, no consideran que pueda hablarse de dependencia entre variables si a un valor de una de ellas puede corresponder más de un valor de la otra. *Influye el ser fumador pero tampoco mucho ya que hay personas fumadoras sin dichos trastornos y no fumadoras con ellos.*

En las nubes de puntos, alumnos que comparan globalmente la gráfica basan, sin embargo, el juicio en ejemplos concretos que no cumplen el tipo de relación esperado, no admitiendo la dispersión en la nube de puntos. *Porque, por ejemplo, hay un grupo que tiene 5 partidos empatados y está en 16 y 20 posición cada uno y si dependiera de alguna forma los partidos empatados estaría en igual posición».*

Otras veces (ejemplos aislados de alumnos) se exige incluso la existencia de una expresión algebraica que ligue las variables, para las nubes de punto. Este caso es aun más restringido, dentro de la concepción determinista o funcional de la asociación, puesto que hay funciones no expresables algebraicamente: *Es un caso más o menos natural, no depende de los empates el lugar de la clasificación. No se sigue una regla que diga que haya relación.* El mismo alumno en el ítem 7 dice: *En la gráfica se observa que a mayor consumo diario de proteínas hay menor natalidad. Pienso que en esta gráfica sí se sigue una regla, aunque no me han enseñado a averiguarla.*

En la comparación de muestras relacionadas, algunos alumnos esperan que la variación de la variable dependiente sea del mismo signo o incluso de la misma magnitud en todos los casos: *Los he comparado y he comprobado que después del tratamiento en algunas mujeres disminuye, pero que en otras aumenta, dependiendo de cada mujer.*

Concepción unidireccional

A veces no se admite la asociación de tipo inverso, teniéndose en cuenta la intensidad, pero no el signo de la asociación, e incluso llegando a considerar la dependencia inversa como independencia. En particular, este fenómeno se observa en el ítem 2, en el que un número importante de sujetos coinciden en manifestar que no existe dependencia, indicando, por tanto, este tipo de concepción: *Según estos datos, creo que padecer trastornos no se debe a la dieta empleada, ya que al usar la blanda sólo los tuvieron 7 ancianos de los 15 que la utilizaron y de los que no la emplearon 9 tuvieron trastornos y 1 no los tuvo.*

Asimismo, en el juicio de asociación emitido respecto a este ítem, tanto en la muestra piloto, como en la de comparación, se observa un importante número de alumnos que ha declarado la independencia, frente a los que lo han hecho en el ítem 3 correspondiente a dependencia directa. Esporádicamente se presenta también en otros ítems.

En el ítem 1, un alumno de la muestra piloto, que ha empleado una sola distribución condicional, deduce la independencia, que es la solución correcta, pero usando este argumento que es incorrecto: *Parece ser que no es o no depende mucho el fumar para padecer trastornos bronquiales, ya que entre los que no fuman y no padecen trastornos bronquiales hay un número más pequeño del que hay en el casillero de los fumadores (60) y, si el no*

fumar dependiera de alguna manera el número de personas (en los no fumadores), debería superar en gran porcentaje a los que fuman.

Concepción localista

Con frecuencia los alumnos basan su juicio en el estudio de casos aislados. Si en ellos se confirma un cierto tipo de asociación, la adoptan como juicio. En el caso de las tablas de contingencia, un número importante de sujetos no considera el problema como de comparación de probabilidades, teniendo sólo en cuenta los casos que confirman la asociación o bien una sola distribución condicional e incluso sólo la celda de máxima frecuencia: *Depende en gran parte porque la mayoría de los que padecen trastornos bronquiales fuman.*

En los diagramas de dispersión se seleccionan puntos que favorecen un cierto tipo de asociación, tomándolos como argumento del tipo de juicio dado, sin tener en cuenta la representatividad de los mismos respecto al resto de los puntos, esto es, sin considerar la información relativa al consenso (Poza 1987). Todos estos casos pueden también ser englobados en la teoría de Kahneman y Tversky (1973), que indica que los sujetos no consideran la importancia de la tasa global de ocurrencia de un caso o casos en una población o grupo al efectuar una predicción sobre la relación entre variables. *Ejemplo: el juez x ha dado el número 1 a la que él cree que es más guapa, sin embargo el juez y sitúa a esa misma muchacha en el puesto 4».*

CONCLUSIONES

En este trabajo hemos analizado los juicios de asociación y estrategias empleadas en los mismos por alumnos del curso de orientación universitaria, previo a la ins-

trucción en el tema. El análisis realizado muestra, por un lado, la riqueza de conceptos y teoremas en acto que el alumno emplea de forma intuitiva en la resolución de estos problemas. Aunque Vergnaud (1990) reconoce que un teorema en acto no es suficiente, puesto que el alumno ha de ser capaz de reconocer explícitamente su verdad y pertinencia, también indica que *los conceptos y teoremas explícitos no forman más que la parte visible del iceberg de la conceptualización; sin la parte escondida, formada por los invariantes operatorios, esta parte visible no sería nada* (p. 145).

A pesar de que las estrategias correctas y parcialmente correctas, y los elementos conceptuales implícitos detrás de ellas muestran una buena base intuitiva para el aprendizaje de la asociación estadística, hemos mostrado también la existencia de errores y concepciones incorrectas sobre este concepto. La importancia de la identificación de estas concepciones incorrectas es que, como Confrey (1990) indica, son con frecuencia resistentes al cambio. En consecuencia, la planificación de la enseñanza del tema debería organizar situaciones en las que los alumnos se enfrenten con las mismas, con objeto de ayudarles a la superación de sus errores. Consideramos por ello de gran interés continuar la investigación iniciada en este trabajo, con el fin de caracterizar estas diversas concepciones y estudiar su dependencia de variables cognitivas e instruccionales.

AGRADECIMIENTOS

Trabajo parcialmente subvencionado por el proyecto PS90-0286 (DGICYT) y por el Grupo de Investigación: *Grupo Jaén de Didáctica*. Código 1349. Junta de Andalucía.

REFERENCIAS BIBLIOGRÁFICAS

- ALLAN L. G. y JENKINS H. M., 1983. The effect of representations of binary variables on judgment of influence, *Learning and Motivation*, Vol. 14, pp. 381-405.
- ALLOY, L.B. y TABACHNIK, N., 1984. Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information, *Psychological Review*, Vol. 91, pp. 112-149.
- ARTIGUE, M., 1990. Épistémologie et Didactique, *Recherches en Didactique des Mathématiques*, Vol. 10(2-3), pp. 241-286.
- BATANERO, C., ESTEPA, A., GODINO, J. D. y GREEN, D.R. (en prensa). Judgments of association in contingency tables, *Journal for Research in Mathematics Education*.

- BARBANCHO A. G., 1973. *Estadística elemental moderna*. (Ariel: Barcelona) (Cuarta edición, reimpresión de 1975).
- BRENNAN, R. L., 1983. *Elements of Generalizability Theory*. (ACT Publications: Iowa).
- CHAPMAN, L.J. y CHAPMAN, J.P., 1969. Illusory correlation as an obstacle to the use of valid Psychodiagnostic signs, *Journal of Abnormal Psychology*, Vol. 74, pp. 271-280.
- CONFREY, J., 1990. A review of the research on student conceptions in Mathematics, Science, and Programming, *Review of Research in Education*, Vol. 16(1), pp. 3-55.
- CROCKER, J., 1981. Judgment of covariation by social perceivers, *Psychological Bulletin*, Vol. 90, 2, pp. 272-292.
- CUADRAS, C. M., 1991. *Métodos de análisis multivariante*. (PPU: Barcelona).
- ESTEPA, A., 1994. *Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores*. Tesis doctoral. Departamento de Didáctica de la Matemática. Universidad de Granada. Edición en microfilm. Barcelona: ETD Micropublicaciones, S. L.
- HILDEBRAND, D. K., LANG, J.D. y ROSENTHAL, H., 1977. *Analysis of nominal data*. (Sage University Paper: Londres).
- INHELDER, B. y PIAGET, J., 1955. *De la logique de l'enfant à la logique de l'adolescent*. (Presses Universitaires de France: París). (Trad. cast., primera reimpresión, 1985. Paidós: Barcelona).
- JENKINS, H.M. y WARD, W.C., 1965. Judgment of the contingency between responses and outcomes, *Psychological Monographs*, Vol. 79, pp. 1-17.
- JENNINGS, D. L., AMABILE, T. M. y ROSS, L., 1982. Informal covariation assessment: Data-based versus theory-based judgments, en D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*, pp. 211-230. (Cambridge University Press: Nueva York).
- KAHNEMAN, D. y TVERSKY, A., 1973. On the psychology of prediction, en D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*, 1982, pp. 48-68. (Cambridge University Press: Nueva York).
- KAHNEMAN, D. y TVERSKY, A., 1982. On the psychology of prediction, en D. Kahneman, P. Slovic y A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases*, pp. 48-68. (Cambridge University Press: Nueva York).
- KILPATRICK, J., 1978. Variables and methodologies in research on problem solving, en L.L. Hatfield y D.A. Bradboard (eds.), *Mathematical problem solving: papers from a research workshop*, pp. 7-20. (Eric/Smeac: Columbus, Ohio).
- PEARSON, K., 1920. Notes on the history of correlation. *Biometrika*. Vol. 13, pp. 25-45, en Pearson, E.S. y Kendall, M., (eds.), 1970, *Studies in the history of Statistics and probability*, Vol. 1, pp. 185-205. (Charles Griffin: Londres).
- PÉREZ ECHEVERRÍA, M. P., 1990. *Psicología del razonamiento probabilístico*. (Ediciones de la Universidad Autónoma de Madrid: Madrid).
- POZO, J.I., 1987. *Aprendizaje de la ciencia y pensamiento causal*. (Visor: Madrid).
- SHAKLEE, H. y MINS, M., 1982. Sources of error in judging event covariations: Effects of memory demands, *Journal of Experimental Psychology Learning, Memory and Cognition*, Vol. 8(3), pp. 208-224.
- SHAKLEE, H. y TUCKER, D., 1980. A rule analysis judgments of covariation between events, *Memory and Cognition*, Vol. 8, pp. 459-467.
- SMEDLUND, J., 1963. The concept of correlation in adults, *Scandinavian Journal of Psychology*, Vol. 4, pp. 165-174.
- VERGNAUD, G., 1990. La théorie des champs conceptuels, *Recherches en Didactique des Mathématiques*, Vol. 10(2-3), pp. 133-170.
- WRIGHT, J.C. y MURPHY, G.L., 1984. The utility of theories in intuitive statistics: the robustness of theory-based judgments, *Journal of Experimental Psychology General*, Vol. 113(2), pp. 301-322.

APÉNDICE

Ítem 1. En un centro médico se han observado a 250 personas para estudiar si el hábito de fumar tiene alguna relación con los trastornos bronquiales, obteniéndose los siguientes resultados.

| | Fuma | No fuma | Total |
|----------------------------------|------|---------|-------|
| Padece trastornos bronquiales | 90 | 60 | 150 |
| No padece trastornos bronquiales | 60 | 40 | 100 |
| Total | 150 | 100 | 250 |

Utilizando estos datos *razona* si padecer o no trastornos bronquiales depende de ser fumador o no. Indica cómo has usado los números de la tabla.

Ítem 2. Se quiere estudiar si la ingestión de una dieta blanda incide en que los ancianos padezcan trastornos digestivos. Para ello se han observado a 25 ancianos, obteniéndose los siguientes resultados.

| | Tuvo trastornos digestivos | No tuvo trastornos digestivos |
|----------------------------------|----------------------------|-------------------------------|
| Padece trastornos bronquiales | 7 | 8 |
| No padece trastornos bronquiales | 9 | 1 |

Utilizando estos datos *razona* si padecer o no trastornos digestivos depende o no de ingerir una dieta blanda en estos ancianos. Indica cómo has usado los números de la tabla.

Ítem 3. Se quiere estudiar si la vida sedentaria (llevar una vida sin realizar ningún tipo de ejercicio físico) puede influir en la aparición de alergia en la piel. Para ello se han observado 30 personas obteniendo los resultados siguientes:

| | Padece alergia en la piel | No padece alergia en la piel |
|-----------------------------|---------------------------|------------------------------|
| Forma de vida sedentaria | 13 | 3 |
| Forma de vida no sedentaria | 2 | 12 |

Utilizando estos datos *razona* si padecer alergia en la piel depende o no de llevar una forma de vida sedentaria. Indica cómo has usado los números de la tabla.

Ítem 4. La siguiente tabla nos da el número de estudiantes que aprobaron o suspendieron un examen teniendo en cuenta el tiempo que cada estudiante dedicó a prepararlo.

| | Tiempo dedicado a preparar el examen | | | Total |
|-----------|--------------------------------------|--------------------|-----------------|-------|
| | Menos de 5 horas | Entre 5 y 10 horas | Más de 10 horas | |
| Aprobados | 11 | 25 | 35 | 71 |
| Suspensos | 15 | 7 | 7 | 29 |
| Total | 26 | 32 | 42 | 100 |

INVESTIGACIÓN Y EXPERIENCIAS DIDÁCTICAS

Utilizando estos datos *razona* si aprobar o suspender el examen depende del tiempo dedicado a prepararlo. Indica cómo has usado los números de la tabla.

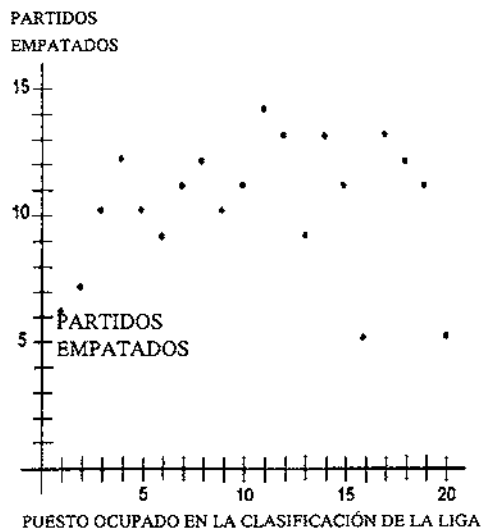
Ítem 5. En un experimento se han obtenido los siguientes datos sobre la lateralidad de las manos y la lateralidad del ojo referidos a personas.

| | Ojo izquierdo | Ambiocular | Ojo derecho |
|-------------|---------------|------------|-------------|
| Zurdos | 3 | 6 | 2 |
| Ambidextros | 2 | 2 | 2 |
| Diestros | 5 | 9 | 4 |

Utilizando estos datos *razona* si la lateralidad del ojo depende de la lateralidad de las manos. Indica cómo has usado los números de la tabla.

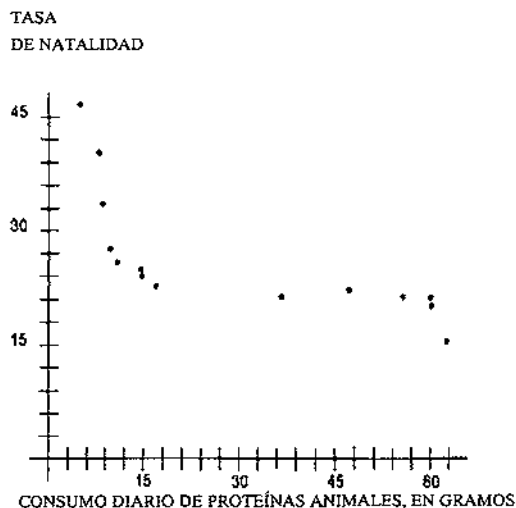
Ítem 6. En el siguiente diagrama hemos representado el puesto ocupado por cada uno de los 20 equipos de primera división de la liga de fútbol en la temporada 1987-88 y los partidos empatados.

A la vista del gráfico, ¿crees que la relación entre el lugar ocupado en la clasificación final de la liga y el número de partidos empatados es: *directa*, *inversa*, *ninguna relación*? (Elige una de estas opciones explicando la respuesta.)



Ítem 7. En un estudio sociológico se han recogido datos relativos a la tasa de natalidad y consumo diario de proteínas animales en diferentes países representándolos en el siguiente gráfico:

A la vista del gráfico, ¿crees que la relación entre el consumo diario de proteínas animales y la tasa de natalidad es: *directa*, *inversa*, *ninguna relación*? (Elige una de estas opciones explicando la respuesta.)

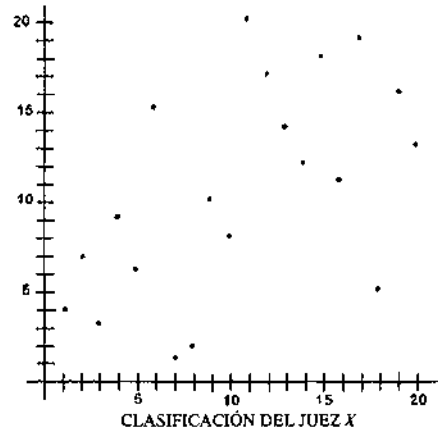


INVESTIGACIÓN Y EXPERIENCIAS DIDÁCTICAS

Ítem 8. En un concurso de belleza clasifican a las 20 concursantes dos jueces (juez X, juez Y) de acuerdo con sus preferencias. El puesto obtenido por cada una en la clasificación de cada juez lo hemos representado en el siguiente gráfico:

¿Crees que la relación entre el lugar asignado a cada concursante por cada juez es: *directa*, *inversa*, *ninguna relación*? (Elige una de estas opciones explicando la respuesta.)

CLASIFICACIÓN DEL JUEZ Y



Ítem 9. Al medir la presión sanguínea antes y después de haber efectuado un cierto tratamiento médico a un grupo de 10 mujeres se obtuvieron los valores siguientes:

Presión sanguínea en cada mujer

| Mujer | Sra. A | Sra. B | Sra. C | Sra. D | Sra. E | Sra. F | Sra. G | Sra. H | Sra. I | Sra. J |
|-------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Antes del tratamiento | 115 | 112 | 107 | 119 | 115 | 138 | 126 | 105 | 104 | 115 |
| Después del tratamiento | 128 | 115 | 106 | 128 | 122 | 145 | 132 | 109 | 102 | 117 |

Utilizando estos datos, responde: ¿Depende la presión sanguínea de haber tomado o no el tratamiento?

¿Por qué? (Indica cómo has usado los números de la tabla.)

Ítem 10. Al medir el nivel de azúcar en la sangre en escolares de ambos sexos se obtuvieron los datos siguientes:

Nivel de azúcar en la sangre

| Alumno | A | B | C | D | E | F | G | I | J | K | L | M | N | Ñ | O | P | Q | R | S | T |
|------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sexo | V | V | V | V | V | V | V | V | V | V | H | H | H | H | H | H | H | H | H | H |
| Nivel de azúcar en la sangre | 9 | 0 | 9 | 8 | 6 | 7 | 4 | 9 | 8 | 9 | 6 | 0 | 7 | 0 | 8 | 3 | 6 | 7 | 7 | 3 |

V = varón, H = hembra.

Utilizando estos datos, responde: ¿Depende el nivel de azúcar en la sangre del sexo de los alumnos?

¿Por qué? (Indica cómo has usado los números de la tabla.)