

**EFFECTO DE LOS FORMATOS Y EL TIPO DE INFORMACIÓN SOBRE LAS
RESPUESTAS Y ESTRATEGIAS PARA RESOLVER UN PROBLEMA BINARIO
DE PROBABILIDAD CONDICIONAL**

**ANA MAYERLY RÁTIVA HERNÁNDEZ
DIANA PAOLA LOZANO RODRÍGUEZ**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2010**

**EFFECTO DE LOS FORMATOS Y EL TIPO DE INFORMACIÓN SOBRE LAS
RESPUESTAS Y ESTRATEGIAS PARA RESOLVER UN PROBLEMA BINARIO
DE PROBABILIDAD CONDICIONAL**

**ANA MAYERLY RÁTIVA HERNÁNDEZ
DIANA PAOLA LOZANO RODRÍGUEZ**

**Trabajo de Grado para Optar al Título de
Licenciada en Matemáticas**

Director

DR. GABRIEL YÁÑEZ CANAL

Especialidad en Matemática Educativa

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE CIENCIAS
ESCUELA DE MATEMÁTICAS
BUCARAMANGA
2010**

AGRADECIMIENTOS

A Dios porque como siempre nos rodeó de tantas personas maravillosas que en su momento nos fortalecieron, alentaron, corrigieron y guiaron siempre con amor. A ellos que sin nombrarlos ya se saben aludidos, porque estuvieron ahí todo el tiempo y que muchas veces su sola presencia fue suficiente para seguir, que dieron su consejo y abrazo en el momento adecuado y que en algunos momentos sintieron esto como algo propio.

Esta página es corta para nombrarlos a todos y si olvidamos a algunos, ellos entenderán lo limitado de nuestra memoria: al profesor Gabriel Yáñez, a nuestras familias, a todos los profesores y estudiantes de la UIS que participaron en las pruebas, a nuestros amigos y a la iglesia.

A todos gracias, muchas gracias.

Ana Mayerly y Diana Paola.

TABLA DE CONTENIDO

PRESENTACIÓN	16
CAPÍTULO I: ANTECEDENTES	19
1.1. ANTECEDENTES MATEMÁTICOS	20
1.1.1. La probabilidad condicional. Definición.	20
1.1.2. Teorema de Probabilidad Total.	20
1.1.3. Teorema de Bayes.	22
1.1.4. Clasificación de los problemas binarios de probabilidad condicional.	22
1.2. ANTECEDENTES SICOLÓGICO-DIDÁCTICOS	26
1.2.1. Dificultades en la resolución de problemas que involucran el teorema de Bayes	27
1.2.1.1. Díaz & de la Fuente (2006).	27
1.2.1.2. Tarr & Lannin (2005).	28
1.2.1.3. Pollatsek & cols (1987).	30
1.2.1.4. Falk (1986).	31
1.2.2. Formatos de presentación de los problemas de probabilidad condicional.	32
1.2.2.1. Gigerenzer & Hoffrage (1995).	33
1.2.2.2. Martignon & Wassner.	38
1.2.2.3. Lewis & Keren (1999).	41
1.2.2.4. Mellers & McGraw (1999).	45

1.2.2.5. Evans & cols (2000).	47
1.2.3. Lonjedo y Huerta.	48
CAPÍTULO II: METODOLOGÍA	55
2.1. TIPO DE INVESTIGACIÓN Y DESCRIPCIÓN DE LA MUESTRA.	56
2.1.1. Tipo de investigación.	56
2.1.2. Población y muestra.	57
2.2. FASES DE LA INVESTIGACIÓN Y ANÁLISIS DE LOS PROCESOS DE RESOLUCIÓN.	58
2.2.1. Primera fase. <i>Estudio piloto.</i>	58
2.2.2. Segunda fase. <i>Estudio sobre el efecto del formato y el tipo de información.</i>	59
2.2.3. Tercera fase. <i>Estudio sobre las dificultades.</i>	61
CAPÍTULO III: RESULTADOS Y ANÁLISIS	65
3.1. EFECTO DEL FORMATO Y TIPO DE INFORMACIÓN	66
3.2. ANÁLISIS DE LAS DIFICULTADES EN LA SOLUCIÓN DE LOS PROBLEMAS BINARIOS DE PROBABILIDAD CONDICIONAL.	70
3.2.1. Dificultades con el lenguaje utilizado para expresar la condicionalidad.	73
3.2.2. Dificultades con el contexto del problema.	75

3.2.3. Problemas con la definición de probabilidad condicional y sus propiedades.	77
3.2.4. Confusión de la probabilidad conjunta con la condicional y viceversa.	79
3.2.4.1. Confusión entre probabilidad conjunta y condicional.	81
3.2.4.2. Confusión entre probabilidad condicional y conjunta.	83
3.2.5. Sobre las representaciones.	84
3.2.6. Otras.	91
CAPÍTULO IV: CONCLUSIONES	94
REFERENCIAS	100
ANEXOS	105
ANEXO A: PRUEBA PILOTO 1	106
ANEXO B: PRUEBA PILOTO 2	108
ANEXO C: PROBLEMAS INDIVIDUALES	110
ANEXO D: PRUEBA DE ESTUDIO 1	111
ANEXO E: PRUEBA DE ESTUDIO 2	113
ANEXO F: PRUEBA DE UNA ESTUDIANTE	115

LISTA DE DIAGRAMAS

Diagrama 1. Diagrama de árbol para la igualdad (2)	22
Diagrama 2. Diagrama de árbol para un problema binario de probabilidad condicional.	37
Diagrama 3. Versión completa de árbol de frecuencias para problema de mamografía.	39
Diagrama 4. Árbol para las frecuencias de Lewis & Keren (1999) en el problema de la Mamografía.	44
Diagrama 5. Árbol para las frecuencias de Gigerenzer & Hoffrage (1995) en el problema de la Mamografía.	44

LISTA DE TABLAS

Tabla 1. Tabla de probabilidades para la primera igualdad en (2).	21
Tabla 2. Tabla donde se analizan los diferentes niveles de problemas y los procedimientos utilizados.	26
Tabla 3. Combinaciones posibles para problemas de probabilidad condicional, según Lewis & Keren.	41
Tabla 4. Tipos de problemas de probabilidad condicional en la clasificación de Lonjedo & Huerta (2004).	49
Tabla 5. Distribución de los estudiantes de acuerdo a la cantidad de cursos tomados de probabilidad y estadística que presentaron las pruebas de estudio.	57
Tabla 6. Proporción de respuestas correctas, según tipo de información y formato.	66
Tabla 7. Distribución de porcentajes por ítem para las diferentes dificultades.	72
Tabla 8. Distribución de porcentajes de representación por ítem, para las pruebas de estudio 1 y 2.	85

LISTA DE GRÁFICOS

Gráfico 1. Interacción entre formato y tipo de información.	70
Gráfico 2. Diagrama de barras de la representación utilizada para el ítem 1, de los estudiantes que respondieron correctamente.	86
Gráfico 3. Distribución de las representaciones utilizadas en el ítem 2. (Si es respuesta correcta. No es respuesta incorrecta).	86
Gráfico 4. Distribución de las representaciones utilizadas en el ítem 3. (Si es respuesta correcta. No es respuesta incorrecta).	87
Gráfico 5. Distribución de las representaciones utilizadas en el ítem 4. (Si es respuesta correcta. No es respuesta incorrecta).	88
Gráfico 6. Distribución de las representaciones utilizadas en el ítem 5, prueba 1. (Si es respuesta correcta. No es respuesta incorrecta).	89
Gráfico 7. Distribución de las representaciones utilizadas en el ítem 6, prueba 1. (Si es respuesta correcta. No es respuesta incorrecta).	89
Gráfico 8. Distribución de las representaciones utilizadas en el ítem 5, prueba 2. (Si es respuesta correcta. No es respuesta incorrecta).	90
Gráfico 9. Distribución de las representaciones utilizadas en el ítem 6, prueba 2. (Si: respuesta correcta - No: respuesta incorrecta).	90

LISTA DE ANEXOS

ANEXO A: PRUEBA PILOTO 1	106
ANEXO B: PRUEBA PILOTO 2	108
ANEXO C: PROBLEMAS INDIVIDUALES	110
ANEXO D: PRUEBA DE ESTUDIO	111
ANEXO E: PRUEBA DE ESTUDIO 2	113
ANEXO F: PRUEBA DE UNA ESTUDIANTE	115

1. TÍTULO:
EFECTO DE LOS FORMATOS Y EL TIPO DE INFORMACIÓN SOBRE LAS RESPUESTAS Y ESTRATEGIAS PARA RESOLVER UN PROBLEMA BINARIO DE PROBABILIDAD CONDICIONAL*.

2. AUTORAS:
Ana Mayerly Rátiva Hernández
Diana Paola Lozano Rodríguez**

3. PALABRAS CLAVES:
Problema binario de probabilidad condicional
Formato de presentación
Frecuencias naturales
Tipo de información

4. RESUMEN:

Esta investigación estudió el efecto de los formatos y el tipo de información sobre las respuestas y estrategias que utilizan los estudiantes para resolver un problema binario de probabilidad condicional. Para esto se realizaron dos estudios en los que participaron 432 estudiantes de diversas carreras de la Universidad Industrial de Santander (UIS).

En el primer estudio se analizó el efecto de los formatos y el tipo de información sobre las respuestas, cuyos resultados mostraron que para problemas en los que se da información condicional hay efecto tanto del formato como del tipo de información, sin embargo para aquellos que no contienen información condicional el formato no tuvo un efecto significativo.

El segundo estudio trató las dificultades que se presentan a la hora de resolver un problema binario de probabilidad condicional, entre las que se destacaron: confusión de una probabilidad conjunta y una condicional, el uso de las representaciones, el lenguaje utilizado para expresar la condicionalidad, la falacia de la condicional transpuesta.

Los resultados de esta investigación además de tener implicaciones didácticas en el proceso de enseñanza aprendizaje de la probabilidad condicional, permitieron caracterizar las intuiciones y dificultades que presentan las personas a la hora de resolver un problema binario de probabilidad condicional y contrastar este estudio con otros realizados alrededor de este tema.

* Trabajo de Grado

** Facultad de Ciencias. Licenciatura en Matemáticas. Dr. Gabriel Yáñez Canal

- 1. TITLE:**
EFFECT OF FORMATS AND KIND OF INFORMATION ON THE ANSWERS AND STRATEGIES FOR SOLVE A BINARY PROBLEM OF CONDITIONAL PROBABILITY. *

- 2. AUTHORS:**
Ana Mayerly Rátiva Hernández
Diana Paola Lozano Rodríguez **

- 3. KEY WORDS:**
Binary problem of conditional probability
Presentation format
Natural frequencies
Kind of information.

- 4. DESCRIPTION:**

This investigation studied the effect of formats and kind of information on the answers and strategies that use the students for solve a binary problem of conditional probability. For this work we made two researches involving 432 students of different careers of the Industrial University of Santander.

In the first research analyzed the effect of the formats and the kind of information about the answers. The results showed that for problems in which conditional information is given to have effect as conditional formatting and the kind of information, however, for those that do not contain conditional information format had no significant effect.

The second research dealt with the difficulties and confusions that arise when solving a binary problem of conditional probability. Among them stood the confusion of a joint and conditional probability, the misuse of the representations, the language used to express the conditionality and the fallacy of the transposed conditional.

The results of this investigation also have didactics implications in the process of teaching learning of conditional probability, allowed to characterize the insights and difficulties people have in solving a binary problem of conditional probability and contrast this research with others carried out around this topic.

* Grade work.

** Faculty of Science. Licentiate in Mathematics. Dr. Gabriel Yáñez Canal

PRESENTACIÓN

La probabilidad condicional, un tema desconocido para muchos y para quienes lo conocen complicado, es un concepto fundamental de la probabilidad que muestra el efecto que puede tener un evento sobre la ocurrencia de otro. Su comprensión y aprendizaje encierra muchas dificultades porque su cálculo requiere tomar como nuevo espacio muestral el evento condicionante y calcular las probabilidades de los eventos ajustadas a este nuevo espacio. Además, es importante resaltar que generalmente la fundamentación teórica del estudiante sobre este tema no es sólida y el lenguaje en que se acostumbra a presentar este tipo de problemas es un poco confuso, dando paso a dificultades en la comprensión.

Así pues, enfrentarse con éxito a un problema de probabilidad condicional implica más que saber de probabilidades, interpretar todos los datos que se presentan en la parte informativa del problema. Yáñez (2001) presenta una clasificación de los problemas binarios de probabilidad condicional atendiendo a la cantidad de probabilidades condicionales presentes en la parte informativa, analiza cada una de las representaciones utilizadas para resolverlos, encontrando que los diagramas de árbol y las tablas de doble entrada son limitadas para dar una completa solución a todos los tipos de problemas y que la solución algebraica si lo hace.

Por otra parte algunos investigadores en sus trabajos dan cuenta de las dificultades que tienen los estudiantes cuando se enfrentan a un problema de probabilidad condicional mostrando que existen diversas razones para explicar estas dificultades (Assumpta, Díaz & de la Fuente, 2005; Díaz de la Fuente, 2006; Díaz, Ortiz y Serrano, 2007; Falk, 1986). En este sentido Gigerenzer & Hoffrage (1995) realizaron una investigación del razonamiento bayesiano y llegaron a que dicho razonamiento mejora cuando a las personas se les presenta la información

en formato de frecuencias naturales y así mismo argumentaron que dar información matemáticamente equivalente de un problema no implica procedimientos necesariamente equivalentes al resolverlo.

En su investigación Gigerenzer & Hoffrage (1995), tratan un sólo tipo de problema (7iii) según la clasificación realizada por Yáñez (2001), en el cual se encuentra una clara correspondencia entre la información del problema y el diagrama de árbol, lo que produce éxito en la solución del mismo. Hay que decir además que este tipo de problemas son los que comúnmente se trabajan debido a su aplicación en diversos contextos (médicos, psicológicos y judiciales). A pesar de esto es de interés analizar si lo planteado por Gigerenzer & Hoffrage (1995) se aplica a problemas de otro tipo.

De aquí surgieron las siguientes preguntas de investigación: si se cambia el tipo de problema respecto al usado por Gigerenzer & Hoffrage (1995): ¿qué efectos tiene sobre las respuestas el formato de presentación del problema? y estos efectos ¿son simplemente por el formato de presentación y el tipo de problema, o hay otros elementos que influyen? ¿Cuáles?

Para dar respuesta a estos interrogantes, se planteó como objetivo principal estudiar el efecto que tienen los diferentes formatos y el tipo de información que se presenta en un problema binario de probabilidad condicional sobre las respuestas y estrategias que las personas utilizan para resolverlo. Con el fin de dar cumplimiento a éste se elaboraron dos pruebas (pruebas de estudio 1 y 2) en las que se presentaron problemas de probabilidad condicional de diversos tipos y en diferentes formatos, para identificar y analizar las dificultades que presentan los estudiantes al resolver dichos problemas. Adicional a estas pruebas se realizaron otras (problemas individuales) con las que se estudió el efecto del formato y el tipo de información utilizando una prueba de comparación de proporciones y un gráfico de interacción.

Este trabajo expone los resultados de esta investigación y está conformado por cuatro capítulos. En el primer capítulo *“Antecedentes”*, se presenta una síntesis y análisis de los trabajos e investigaciones relacionadas con el tema de estudio.

El segundo capítulo *“Metodología”*, describe aspectos relacionados con el diseño de la investigación, también presenta la descripción de las pruebas y los métodos empleados para el análisis de los datos obtenidos en la aplicación de éstas.

En el tercer capítulo *“Resultados y Análisis”*, se presentan los resultados de la investigación haciendo un análisis cualitativo y cuantitativo de éstos, que permitieron comprobar o refutar las hipótesis planteadas.

En el cuarto capítulo *“Conclusiones”*, se plasman las ideas consideradas como una contribución a la enseñanza y aprendizaje de la probabilidad condicional, resultado del proceso de investigación.

Finalmente se espera que las personas disfruten leyendo el contenido de este trabajo tanto como sus autoras disfrutaron escribiéndolo. Y que éste genere la motivación para continuar con la ampliación y aplicación de los temas tratados con el fin de mejorar los procesos de enseñanza-aprendizaje en la resolución de problemas de probabilidad condicional.

CAPÍTULO I

ANTECEDENTES

Este capítulo presenta la base teórica que fundamenta esta investigación. Como la investigación tiene un componente matemático (problemas de probabilidad condicional y clasificación) y uno didáctico (formato de presentación de los problemas y dificultades en la resolución de los problemas de probabilidad condicional), el capítulo está dividido en dos secciones: la primera, llamada “Antecedentes matemáticos”, trata definiciones, propiedades y teoremas importantes relacionados con la probabilidad condicional; la segunda, llamada “Antecedentes psicológico-didácticos” muestra los resultados de investigaciones sobre problemas de probabilidad condicional y su implicación en la enseñanza.

1.1. ANTECEDENTES MATEMÁTICOS

En esta sección se presenta la teoría formal de la probabilidad condicional desde su definición hasta el teorema de Bayes. Además se muestra y se describe la clasificación de los problemas binarios de probabilidad condicional realizada por Yáñez (2001) que es otro de los fundamentos teóricos importantes de esta investigación.

1.1.1. La probabilidad condicional. Definición. Sean A y B eventos y $P(B) > 0$, la probabilidad condicional de A dado B está definida como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

Y se lee “la probabilidad de A dado B ”, donde B es el evento condicionante y A es el evento condicionado.

1.1.2. Teorema de Probabilidad Total. Sean A_1, A_2, \dots, A_n sucesos mutuamente excluyentes y exhaustivos y B cualquier evento, entonces

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B | A_i) P(A_i) \quad (2)$$

Estas igualdades dan lugar a dos tipos de representaciones gráficas que facilitan la solución del problema. “Para una mejor comprensión y sin pérdida de generalidad, se consideran sólo dos eventos A y \bar{A} . Considerando que los eventos B y \bar{B} constituyen también una partición, se obtiene una expresión similar a la anterior para cada uno de los eventos A y \bar{A} . La integración de la primera igualdad en (2) para cada uno de los eventos en juego, permite construir la [Tabla 1] de probabilidades.

Por otra parte usando las dos igualdades en (2) tanto para B como para \bar{B} se obtiene el diagrama de árbol [Diagrama 1] que al leerlo de izquierda a derecha expresa la igualdad (2) leída de derecha a izquierda”. (Yáñez, 2003)

	B	\bar{B}	Marginal
A	$P(A \cap B)$	$P(A \cap \bar{B})$	$P(A)$
\bar{A}	$P(\bar{A} \cap B)$	$P(\bar{A} \cap \bar{B})$	$P(\bar{A})$
Marginal	$P(B)$	$P(\bar{B})$	1

Tabla 1. Tabla de probabilidades para la primera igualdad en (2)

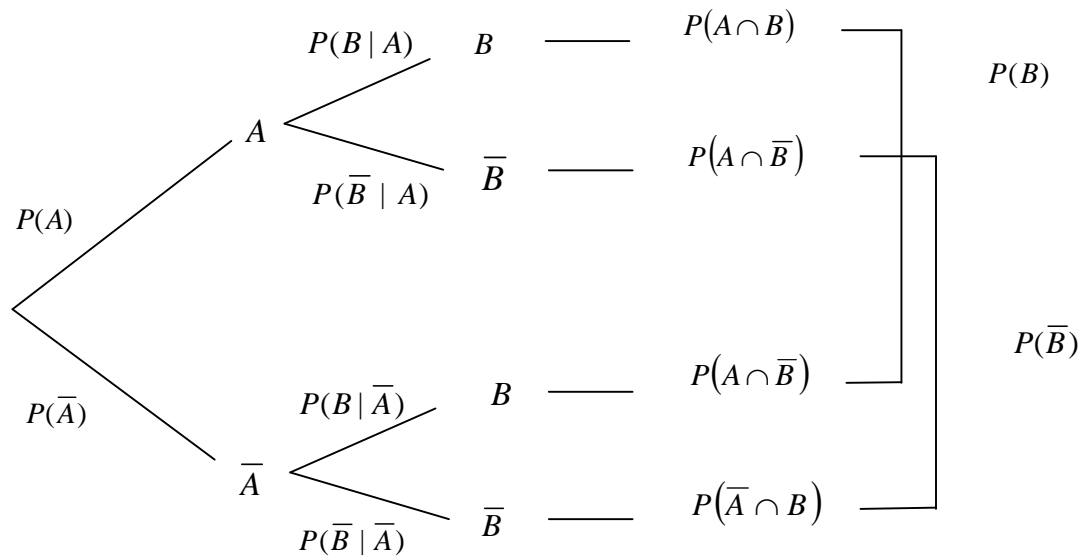


Diagrama 1. Diagrama de árbol para la igualdad (2)

1.1.3. Teorema de Bayes. Sean A_1, A_2, \dots, A_n sucesos mutuamente excluyentes y exhaustivos con $P(A_i) > 0$ para cada A_i , y sea B cualquier evento de probabilidad no nula, entonces:

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{i=1}^n P(B | A_i)P(A_i)} \quad (3)$$

1.1.4. Clasificación de los problemas binarios de probabilidad condicional.

A continuación se presenta una clasificación de los problemas binarios de probabilidad condicional dada por Yáñez (2001). Sean A y B dos eventos con sus respectivos complementos denotados por \bar{A} y \bar{B} ; sean, además, las probabilidades condicionales $P(A|B)$, las intersecciones, $P(A \cap B)$, y las probabilidades marginales $P(A)$.

Tomando los eventos A y B con sus complementos, las probabilidades que se pueden construir con ellos son las siguientes:

$P(A \cap B)$	$P(A B)$	$P(B A)$	$P(B)$
$P(A \cap \bar{B})$	$P(A \bar{B})$	$P(\bar{B} A)$	$P(\bar{B})$
$P(\bar{A} \cap B)$	$P(\bar{A} B)$	$P(B \bar{A})$	$P(A)$
$P(\bar{A} \cap \bar{B})$	$P(\bar{A} \bar{B})$	$P(\bar{B} \bar{A})$	$P(\bar{A})$

Es decir, 8 condicionales, 4 intersecciones y 4 marginales. Encontrar estos 16 elementos es lo que Yáñez (2001) denomina resolver completamente un problema de probabilidad condicional.

Ahora bien, como es claro que para resolver completamente un problema de probabilidad condicional no se requiere contar con todos los 16 elementos, se plantea la pregunta de cuál es el número mínimo que se requieren para obtener los demás y cuáles pueden ser estos elementos. Para responder estas preguntas se cuenta con las siguientes relaciones básicas de la teoría de probabilidades:

$$P(A) + P(\bar{A}) = 1 \quad (4)$$

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \quad (5)$$

$$P(A | B) + P(\bar{A} | B) = 1 \quad (6)$$

y con las siguientes definiciones:

- Dos marginales, dos intersecciones y dos condicionales que satisfacen (4), (5) y (6) respectivamente, se dice que son *complementarias*.
- Una marginal y una intersección (o dos) que satisfacen (5), se dice que están *asociadas*.

- Una condicional está *asociada* a una marginal cuando su evento condicionante es la marginal considerada.
- Una condicional está *asociada* a dos intersecciones cuando éstas, a su vez, están asociadas y la marginal correspondiente es el evento condicionante de la condicional.

Pues bien, con estas ideas y junto con la definición de probabilidad condicional, Yáñez (2001) concluye que el mínimo número de elementos necesarios para resolver completamente un problema de probabilidad condicional se reduce a conocer tres elementos no complementarios, por ejemplo: $P(A) = x_1$; $P(A \cap B) = y$; $P(A | B) = z_1$.

Tres como número mínimo condujo a Lonjedo & Huerta (2007) a identificar estos problemas como ternarios, notación que no es compartida porque este número responde a la solución completa del problema y no necesariamente en los problemas tradicionales donde se hacen una o dos preguntas. Por esta razón se adoptó la notación problemas binarios para hacer referencia al número de eventos simples que constituyen el problema.

De acuerdo con las posibles formas de escoger los elementos se llega a los tipos de problemas que se clasifican según el número de probabilidades condicionales dadas en la parte informativa del problema.

Así, la clasificación a la que llega Yáñez (2001) es la siguiente:

- **De Nivel 0:** aquellos que en su parte informativa no presentan información condicional.
 - ✓ Caso 1: Tres intersecciones
 - ✓ Caso 2: Dos intersecciones y una marginal
 - ✓ Caso 3: Dos marginales y una intersección

- **De Nivel 1:** aquellos que en su parte informativa presentan una condicional.
 - ✓ Caso 4: Dos marginales y una condicional
 - ✓ Caso 5: Dos intersecciones y una condicional
 - ✓ Caso 6: Una marginal, una intersección y una condicional

- **De Nivel 2:** aquellos que en su parte informativa presentan dos condicionales.
 - ✓ Caso 7: Una marginal y dos condicionales
 De este caso se pueden dar tres tipos de problemas:
 - i) Una condicional asociada con la marginal y la otra asociada con el complemento de la marginal.
 - ii) Una condicional asociada con la marginal dada, y la otra asociada con la marginal.
 - iii) Las condicionales están asociadas con la otra marginal y con su complemento.
 - ✓ Caso 8: Una intersección y dos condicionales

- **De Nivel 3:** aquellos que en su parte informativa presentan tres condicionales.
 - ✓ Caso 9: Tres condicionales

Una vez establecida esta clasificación, Yáñez (2001) analiza hasta qué punto los diagramas de árbol y las tablas permiten resolver estos problemas y de qué manera lo hacen. Los resultados a los que llega se presentan en la Tabla 2 en la que se puede apreciar el nivel del problema y la posibilidad o imposibilidad de resolverlo mediante el diagrama de árbol, las tablas o un análisis algebraico.

Soluciones directas (Direct) en la Tabla 2, hace referencia a las incógnitas cuyo valor está explícito o cuando se tiene que realizar despeje de una ecuación, sustitución simple (S1) cuando al reemplazar el valor de una variable en una

ecuación se obtiene el valor de otra; sustitución doble (S2) cuando se reemplazan dos variables en una ecuación para hallar una tercera; sustitución triple (S3) cuando el número de sustituciones es tres. El signo + significa que tienen un buen comportamiento; es decir, son buenas herramientas para la solución; ± implica que su comportamiento es aceptable mientras que ∓ indica que si bien la solución es posible con ese método, no lo es de manera muy directa, requiere un trabajo mayor. Por último, el símbolo ◦ indica la imposibilidad de resolver el problema mediante dicho diagrama.

Nivel	Caso	Direct.	S1	S2	S3	S2X2	S3X3	Tabla	Árbol
0	3	4	0	0	0	0	0	+	∓
1	6	1	1	2	0	0	0	∓	∓-◦
	7(i)	0	1	2	0	1	0	◦	◦
	7(iii)	0	0	0	1	0	1	◦	◦
3	9	0	0	0	1	0	1	◦	◦

Tabla 2. Tabla donde se analizan los diferentes niveles de problemas y los procedimientos utilizados

1.2. ANTECEDENTES SICOLÓGICO-DIDÁCTICOS

En este apartado se presentan algunos resultados de carácter sicológico-didáctico que están relacionados con los objetivos de esta investigación. Tratar las dificultades que se presentan al resolver problemas de probabilidad condicional constituye un marco de referencia importante a la hora de estudiar el efecto de los formatos y el tipo de información sobre las respuestas y estrategias para resolver un problema de este tipo. Por esto se considera necesario presentar los análisis que sobre estas dificultades realizan Díaz & de la Fuente (2006), Tarr & Lannin (2005), Pollatsek & cols (1987) y Falk (1986). Se presenta también un resumen de las investigaciones realizadas alrededor del efecto de los formatos de presentación de los problemas de probabilidad condicional. Se cierra este

apartado con un resumen de las investigaciones realizadas por Lonjedo y Huerta, quienes plantearon una clasificación de los problemas binarios de probabilidad condicional complementaria a la de Yáñez (2001). Adicionalmente, estos autores exploraron sobre algunos factores que influyen en la solución de un problema de probabilidad condicional, siendo este otro de los puntos a tratar en esta investigación.

1.2.1. Dificultades en la resolución de problemas que involucran el teorema de Bayes.

Con relación a las dificultades en la resolución de los problemas de probabilidad condicional, se presentan algunos estudios Díaz & de la Fuente (2006), Tarr & Lannin (2005), Pollatsek & cols (1987) y Falk (1986), quienes dan a conocer diversas particularidades y generalidades en la forma de razonar de las personas a la hora de resolver problemas que involucran razonamiento condicional.

1.2.1.1. Díaz & de la Fuente (2006). En su estudio *Dificultades en la resolución de problemas que involucran el teorema de Bayes un estudio exploratorio en estudiantes españoles de psicología*, Díaz & de la Fuente (2006) investigaron los procesos llevados a cabo por los estudiantes para la solución de problemas antes y después de la enseñanza de la probabilidad condicional con base en los siguientes aspectos:

- Respuesta correcta
- Identificación correcta de los datos
- No responde
- No identifica datos o identificación incorrecta
- Representación incorrecta.
- Confunde probabilidad condicional y conjunta
- Confunde probabilidad marginal con condicional
- Confunde una probabilidad condicional con su inversa.

- Confunde un suceso con su complementario.
- Error en la partición del espacio muestral.
- Fallo con las operaciones.
- Dificultad en la fórmula de Bayes.
- Error en conceptos y propiedades de la probabilidad.

Las autoras llegaron a la conclusión que los errores se producen en diferentes pasos del proceso de resolución, comenzando por la identificación correcta de los sucesos y sus probabilidades y la correcta partición y subpartición del espacio muestral. A muchos estudiantes les fue difícil diferenciar entre probabilidades simples, compuestas y condicionales, o confundieron una probabilidad condicional con su inversa. El olvido de la fórmula de Bayes también ocasionó algunos errores, pero su número es pequeño en comparación con los causados por identificación de datos y errores de los conceptos que intervienen.

“Otra conclusión es que se necesita más tiempo para enseñar razonamiento bayesiano si se quiere tener éxito con los estudiantes. Ya que es un tema útil en la toma de decisiones, diagnóstico y evaluación y constituye la base del estudio de la inferencia estadística bayesiana y de la correlación” (Díaz & de la Fuente, 2006, p. 91).

1.2.1.2. Tarr & Lannin (2005). Tarr & Lannin (2005), en su estudio *A framework for Assessing Middle School Students' Thinking in Conditional Probability and Independence*, realizaron una investigación en probabilidad condicional e independencia con estudiantes de bachillerato, en la cual demuestran que a través del contexto de situaciones con y sin reemplazo, se puede evaluar el razonamiento de los estudiantes en probabilidad condicional. Para ello desarrollan un marco de referencia que describe y predice el razonamiento de los estudiantes en probabilidad condicional e independencia, considerando que el razonamiento de probabilidad condicional en un estudiante es demostrado por su habilidad al

reconocer y ajustar la probabilidad de un evento cuando éste es cambiado por la ocurrencia de otro evento, y que el razonamiento de independencia se demuestra por la habilidad de reconocer y explicar correctamente cuándo la ocurrencia de un evento no influye en la probabilidad de ocurrencia de otro.

El tipo de investigación trabajado por estos autores fue el estudio de casos (15 estudiantes), donde enfrentaban a cada estudiante a cuatro ítems de independencia y cuatro de probabilidad condicional, de manera alternada, con el fin de explorar el pensamiento a través de cuatro niveles que muestran el desarrollo del pensamiento subjetivo al razonamiento cuantitativo formal.

En particular, el nivel 1 está asociado con el pensamiento subjetivo, el nivel 2 es la etapa de transición entre el pensamiento subjetivo y el pensamiento cuantitativo informal, el nivel 3 es el que involucra uso de pensamiento cuantitativo informal, y el nivel 4 incorpora el razonamiento cuantitativo formal.

Los estudiantes que están en el nivel 1 de pensamiento tienden a emitir juicios subjetivos, generalmente creen que pueden controlar los resultados de un evento e ignoran información cuantitativa relevante cuando formulan juicios de probabilidad.

Los estudiantes que exhiben el nivel 2 algunas veces hacen uso apropiado de información cuantitativa, formulando juicios de probabilidad condicional, pero a menudo se distraen con información irrelevante.

Para el nivel 3 los estudiantes han ganado conocimiento sobre el papel de las cantidades en los juicios de probabilidad condicional. Sin embargo, la mayoría de las veces no asignan probabilidades numéricas, sino frecuencias relativas o algunas proporciones como una estrategia apropiada para determinar probabilidades condicionales en situaciones con o sin reemplazo.

Los estudiantes que muestran nivel 4 en su pensamiento pueden asignar espontáneamente probabilidades numéricas cuando interpretan situaciones de probabilidad y reflejan en sus respuestas cierta sofisticación en el pensamiento probabilístico, pues establecen condiciones para las cuales dos eventos son dependientes o independientes.

Finalmente, los autores concluyen que es pertinente y necesario que en el currículo de matemáticas de la escuela se presenten de la mano estos dos temas (probabilidad condicional e independencia), ya que ambos son conceptos claves en el desarrollo del pensamiento probabilístico. Además, recomiendan la introducción del concepto de independencia a través de la definición de probabilidad condicional, debido a que las representaciones de independencia son un caso especial de probabilidad condicional y es más intuitiva para los estudiantes.

1.2.1.3. Pollatsek & cols (1987). Estos autores realizaron una investigación con estudiantes de psicología que no habían tomado un curso de estadística, a quienes enfrentaron con variados problemas de probabilidad condicional, las conclusiones fueron:

1. Las personas sin instrucción en probabilidad son capaces de comprender el concepto de probabilidad condicional y su dirección.
2. Varias características de los datos sugieren que una de las mayores fuentes de error es la confusión entre las probabilidades condicionales y las probabilidades conjuntas. Se podría especular que algunas personas, en lugar de poseer dos conceptos diferenciados, pueden tener un concepto que es la amalgama de los dos.

3. Aunque se dieron los sesgos causales, no lo fueron de forma dominante. Más aún los datos dan a entender que éstos están muy relacionados con la redacción del problema: se dan más en los problemas redactados en términos de probabilidad que en aquellos redactados en términos porcentuales.
4. Los autores consideran que sus resultados son correspondientes con la hipótesis de que las dificultades con las probabilidades condicionales son a menudo causadas por errores de translación del lenguaje natural al algebraico.
5. Sus datos no reflejaron que los estudiantes hubieran intercambiado $P(A | B)$ y $P(B | A)$, pero sí la tendencia a que la suma de estas probabilidades tiene que ser 100%, tal como también lo relata (Yáñez, 2003, p. 79-80)

1.2.1.4. Falk (1986). En su artículo *Conditional Probabilities: insights and difficultie*, discute tres problemas de las probabilidades condicionales, que deben ser considerados por los profesores. Uno es la confusión de la inversa, también llamada falacia de la condicional transpuesta. Cuando las personas presentan este problema no discriminan adecuadamente entre las dos direcciones de la probabilidad condicional $P(A | B)$ y $P(B | A)$, y una posible explicación dada por el autor de la prevalencia de este error, es que el lenguaje usado en el enunciado de los problemas de probabilidad no es preciso. Cuando se escribe una probabilidad condicional usando la notación matemática, es claro cuál es el suceso condicionante y cuál el condicionado; pero en el lenguaje utilizado en el enunciado, la probabilidad condicional y su inversa no siempre se distinguen claramente entre sí, o de la probabilidad conjunta. Falk observó este error en problemas de contextos médicos, donde se confunde la probabilidad de que una

persona tenga una enfermedad cuando el test diagnóstico le ha dado positivo, con la probabilidad de obtener un test diagnóstico positivo dado que tiene la enfermedad, y mostró cómo este error puede tener consecuencias importantes. Por ejemplo, la confusión entre la probabilidad de que un niño con síndrome de Down obtenga un examen de amniocentesis prenatal positiva, que es alta y el hecho de que, dando la prueba positiva el niño realmente tenga síndrome de Down, que es mucho menor.

Adicionalmente, Falk (1986) trata el problema de la interpretación de la condicionalidad como causalidad que es cuando las personas interpretan la probabilidad condicional $P(A | B)$ como una relación causal implícita, donde el suceso condicionante es la causa y el condicionado es la consecuencia. Y, finalmente, discute sobre la dificultad al definir el evento condicionante, aclarando que éste debe ser definido directamente por el procedimiento experimental del problema, es decir, por el paso a paso de la ocurrencia de los eventos en el problema: *“Aquí no solo importa qué sabemos sino también cómo lo sabemos”* (Falk, 1986, p. 294)

1.2.2. Formatos de presentación de los problemas de probabilidad condicional. Respecto a los formatos de presentación de los problemas de probabilidad condicional y su efecto en los procesos de resolución, existen dos posiciones: Por un lado están aquellos que como Gigerenzer & Hoffrage (1995), Martignon & Wassner (2002), Cosmides & Toby (1996), argumentan que los problemas que se presentan en formato de frecuencias naturales facilitan su comprensión y su solución; opuestos a esta opinión Lewis & Keren (1999), Mellers & McGraw (1999), Evans & cols. (2000), Girotto & González (2000), quienes consideran que no siempre las frecuencias naturales facilitan la resolución del problema.

A continuación se presenta una síntesis de los trabajos más relevantes de cada corriente.

1.2.2.1. Gigerenzer & Hoffrage (1995). En su investigación titulada: *How to improve bayesian reasoning without instruction: frequency formats*, Gigerenzer & Hoffrage (1995) presentan el resultado de un análisis teórico y empírico en el que concluyen que el razonamiento bayesiano se mejora si la información que se suministra en el problema está en formato de frecuencias. Los autores conjeturaron que dar información matemáticamente equivalente de un problema no implica procedimientos necesariamente equivalentes al resolverlo.

Para comprobar su hipótesis diseñaron dos estudios en los que enfrentaban a un grupo de 75 estudiantes voluntarios de la universidad de Salzburgo, Austria y Konstanz, Alemania, con 15 problemas de probabilidad condicional redactados en diferentes formatos y con diferente menú de información, en los que se trata de hallar una probabilidad condicional inversa que exige utilizar el teorema de Bayes. Tomaron tres formatos: el de frecuencias absolutas, probabilidad, y frecuencias relativas. Con menú de información se hace referencia a la información dada y los clasifican en menú estándar, que da como información $P(H)$, $P(D|H)$ y $P(D|\bar{H})$, y el menú corto que suministra $P(D \cap H)$ y $P(D \cap \bar{H})$ o $P(D \cap H)$ y $P(D)$, donde H (o \bar{H} su negación) es el evento del que se quiere conocer su probabilidad dado el evento D . Combinando los formatos con los menús se obtienen cuatro versiones de un problema, que se ejemplifican de la siguiente forma:

- **Formato estándar de probabilidad:** La probabilidad de cáncer de pecho es 1% para las mujeres de 40 años que participaron en estudios de rutina. Si una mujer tiene cáncer de pecho la probabilidad de que obtenga una mamografía positiva es de 80%. Si una mujer no tiene cáncer de pecho la probabilidad de que también obtenga una mamografía positiva es de

9.6%. Una mujer en este grupo de edad obtuvo una mamografía positiva en un estudio de rutina. ¿Cuál es la probabilidad de que realmente tenga cáncer de pecho?

- **Formato estándar de frecuencia:** 10 de cada 1000 mujeres de 40 años de edad que participaron en estudios de rutina tienen cáncer de pecho. 8 de cada 10 mujeres con cáncer de pecho obtendrán una mamografía positiva. 95 de cada 990 mujeres sin cáncer de pecho también obtendrán una mamografía positiva. Se tiene aquí una muestra nueva de mujeres de 40 años que obtuvieron una mamografía positiva en estudios de rutina. ¿cuántas de estas mujeres espera que tengan realmente cáncer de pecho?
- **Formato corto de probabilidad:** la probabilidad de que una mujer de 40 años obtenga una mamografía positiva en un estudio de rutina es 10.3%. La probabilidad de cáncer de pecho y mamografía positiva es 0.8% para una mujer de 40 años que participa en un estudio de rutina. Una mujer de este grupo de edad tuvo una mamografía positiva en un estudio rutinario, ¿cuál es la probabilidad de que realmente tenga cáncer de pecho?
- **Formato corto de frecuencia:** 103 de cada 1000 mujeres de 40 años de edad obtienen una mamografía positiva en un estudio rutinario. 8 de cada 1000 mujeres de 40 años que participaron en estudios de rutina tienen cáncer de pecho y una mamografía positiva. Se tiene una muestra representativa de mujeres de 40 años que obtuvieron una mamografía positiva en estudios rutinarios. ¿cuántas de estas mujeres esperarías que realmente tengan cáncer de pecho?
- **Formato en frecuencias relativas:** 1% de las mujeres de 40 años que participaron en pruebas de rutina tiene cáncer de pecho. 80% de las mujeres con cáncer de pecho obtendrá mamografías positivas. 9.6% de las

mujeres sin cáncer de pecho también obtendrá mamografías positivas. Una mujer en este grupo de edad tuvo un resultado positivo en una prueba de rutina, ¿Cuál es la probabilidad de que realmente tenga cáncer de pecho?

Las conclusiones a las que llegaron los autores fueron:

1. Los formatos de frecuencia producen una proporción más alta de algoritmos bayesianos¹ que los formatos de probabilidad y esto se debe a que los formatos de frecuencia corresponden a la manera secuencial en la que se adquiere la información en el muestreo natural.
2. El formato de probabilidad para el menú corto produce mayor porcentaje de respuestas correctas (28%) que para el menú estándar de probabilidad cuyo porcentaje fue 16%. Esto se debe a que para resolver problemas de probabilidad condicional presentados en formato corto, no se requiere la utilización de la fórmula de Bayes, sino únicamente la definición de probabilidad condicional.
3. Los formatos de frecuencia producen la misma proporción de algoritmos bayesianos para los dos menús.
4. Los formatos de frecuencia relativa producen la misma proporción (pequeña) de algoritmos bayesianos que los formatos de probabilidad. El cálculo de algoritmos bayesianos es equivalente en ambos formatos

¹Un algoritmo bayesiano es un proceso inferencial en donde:1) la probabilidad o la frecuencia estimada es exactamente la misma que el valor calculado al aplicar el teorema de Bayes 2) el proceso de solución muestra el uso de ecuaciones equivalentes al teorema de Bayes. (Gigerenzer & Hoffrage, 1995, p. 26)

Para finalizar, los autores plantean algunas consecuencias prácticas de este estudio. La primera es que los procedimientos utilizados para dar respuesta a un problema de probabilidad condicional no pueden estar separados de la información sobre la cual operan y de la manera como se presenta. También recomiendan enseñar a la gente a traducir probabilidades en frecuencias, en lugar de enseñar a insertar probabilidades en la fórmula de Bayes.

En este punto es pertinente notar que los problemas presentados en formato estándar de probabilidad por estos autores corresponden al caso 7(iii) en la clasificación de Yáñez (2001) y conducen directamente a la construcción de un diagrama en árbol y a su solución cuando se conoce la forma de operar con esta representación. En el caso de los formatos cortos, las respuestas se obtienen directamente cuando se comprende el concepto de probabilidad condicional. Esta aclaración obliga a preguntarse:

1. ¿cómo se comportan las frecuencias si se cambia la pregunta del problema? Es decir, si en lugar de indagar por una probabilidad condicional se pregunta por una intersección o una marginal. ¿Las frecuencias naturales siguen siendo mejores que las probabilidades?
2. ¿Qué sucede con problemas de otro tipo, aquellos para los que la información no conduce a una construcción directa de diagrama en árbol o una tabla de doble entrada?

Precisamente esta investigación busca dar respuesta a estas preguntas. Se empieza haciendo un análisis teórico de la primera. En el Diagrama 2, las probabilidades subrayadas en gris indican los datos que se dan como información (en los problemas de Gigerenzer & Hoffrage, 1995), con éstos es posible hallar $P(\bar{B})$, $P(\bar{A}|B)$ y $P(\bar{A}|\bar{B})$ utilizando las ecuaciones (4) y (6). Al multiplicar las marginales por las respectivas condicionales en cada rama, se obtienen las

intersecciones $P(A \cap B)$, $P(A \cap \bar{B})$, $P(\bar{A} \cap B)$ y $P(\bar{A} \cap \bar{B})$ y la suma de las intersecciones adecuadas, como lo indica el mismo diagrama, arroja como resultado las marginales $P(A)$ y $P(\bar{A})$. Con estos datos ya es posible responder a cualquier pregunta que se haga- claro está que hay unas más fáciles de responder que otras-, haciendo un diagrama correcto y sabiendo aplicar las relaciones entre probabilidades.

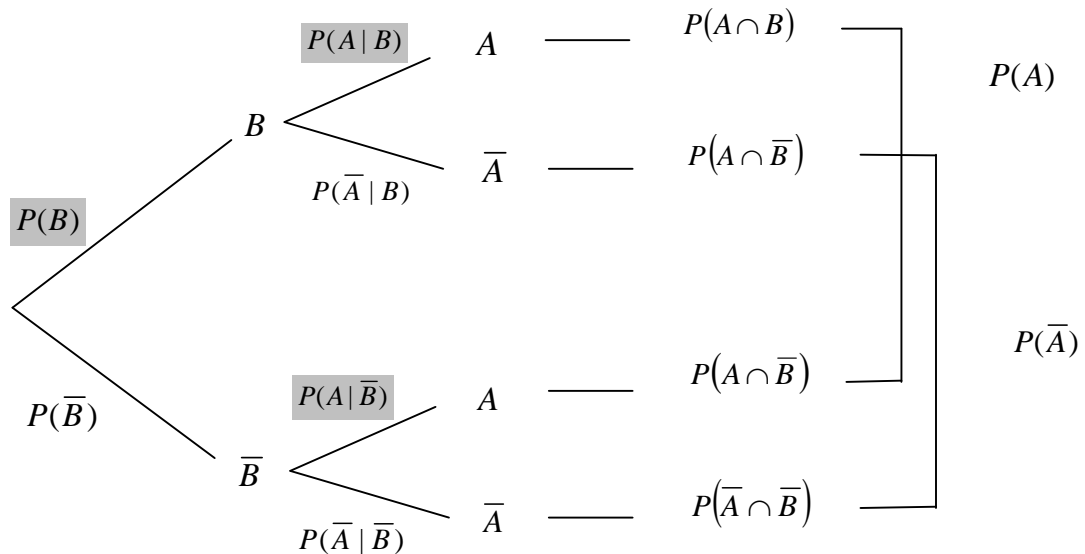


Diagrama 2. Diagrama de árbol para un problema binario de probabilidad condicional.

Ahora se verá lo que sucede cuando se cambia el tipo de información y se pide responder a la misma pregunta. Como este trabajo está enfocado a los casos 2 y 5, según la clasificación de Yáñez (2001), se hace un análisis *a priori* de lo que se esperaría. La idea es presentar información no asociada directamente con la fórmula de Bayes y preguntar por $P(B | A)$.

El caso 7 lleva a hacer una construcción en secuencia y siempre hacia adelante del diagrama de árbol, mientras que para el caso 2 y 5 la construcción implica ir hacia adelante y hacia atrás para completar todos los datos, y esto de alguna

manera complica la obtención de los datos necesarios para dar solución al problema.

Sin embargo para el caso 2, por el tipo de información que se presenta, los datos se pueden ajustar mucho mejor a una tabla de doble entrada, lo que facilitaría mucho la obtención de los demás datos y, por ello, la solución del problema. Pero para el caso 5 no se puede decir que la tabla o el árbol se comporte uno mejor que otro. Luego se podría esperar que los problemas del caso 2 sean más fáciles de resolver que los de caso 5, siempre y cuando se use un diagrama adecuado.

De este análisis se podría inferir *a priori* que cambiar el tipo de información hace que unos problemas sean más fáciles de resolver que otros y que, por consiguiente, resolver un problema de probabilidad condicional con éxito está ligado al tipo de información que se presenta.

1.2.2.2. Martignon & Wassner. En el artículo *Teaching decision making and statistical thinking with natural frequencies*, Martignon & Wassner (2002), desde un análisis teórico y experimental, mostraron que los formatos de frecuencias naturales incrementan el buen desempeño en la resolución de los problemas de probabilidad condicional que involucran el teorema de Bayes.

Uno de los argumentos a favor de las frecuencias naturales es que éstas se ajustan a la manera en que recibimos la información del entorno; además que permiten una representación adecuada de los datos y, a su vez, llegar a los cálculos y procedimientos para solucionar el problema. El término “representación” se aplica a los procesos y productos que son observables externamente.

Los humanos, en particular los estudiantes, tienen intuiciones primarias respecto a la probabilidad condicional. Los autores creen que estas intuiciones son el resultado de elementos genéticos y procesos de aprendizaje e interacción con el

entorno. La mente trabaja con representaciones del entorno para el cual se ha adaptado. Cosmides & Toby (1996), agregan que la mente se adapta a la representación por frecuencias naturales, un camino que está basado en un conteo simple y es el mismo usado por nuestros ancestros cuando ellos hacían inferencia bajo incertidumbre. Una recomendación didáctica es el uso de las frecuencias naturales tan a menudo como sea posible, para favorecer la visión de los estudiantes en problemas de probabilidad condicional.

El proceso de resolución comienza identificando los sucesos a que se refiere la pregunta del problema y la asignación de una notación. A continuación, se usa un esquema como el del Diagrama 3, para representar los datos del problema de la mamografía y como ayuda en su resolución.

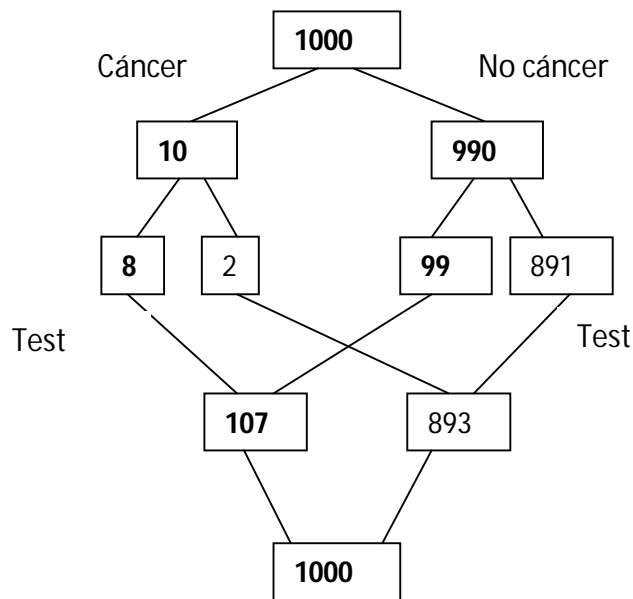


Diagrama 3. Versión completa de árbol de frecuencias para problema de mamografía.

Los autores señalan que el número de respuestas correctas incrementa debido a la estrecha relación entre esta representación y la manera inductiva en la que se procesa la información en las tareas bayesianas. Para construir el árbol, primero, se divide la muestra (1000) en función de la tasa base y se obtienen así dos grupos: cáncer (10) y no cáncer (990). La división en el primer nivel del árbol produce una división binaria. En seguida, se incluye la información condicional y se llega al tercer nivel del árbol, que consiste en una segmentación que produce cuatro sucesos intersección. A partir de esto se suman los sucesos que corresponden a la condición (positivo) y se obtiene la cuarta rama con el total de casos positivos (107). En este momento ya se puede resolver el problema de una manera muy sencilla, simplemente aplicando la regla de Laplace (casos favorables sobre casos posibles), y el resultado es el mismo que si se aplicara la fórmula de Bayes.

Otra ventaja de esta representación es que se puede hacer cualquier “inversión” Bayesiana con la información dada (es decir, de abajo hacia arriba), lo que permite obtener, por ejemplo, la especificidad (probabilidad de obtener mamografía positiva teniendo cáncer), y la sensibilidad (probabilidad de obtener mamografía negativa no teniendo cáncer) del test.

En un experimento con estudiantes de secundaria los autores compararon el desempeño de aquellos que habían sido enseñados con árboles de frecuencias y aquellos que fueron enseñados con probabilidad de manera clásica. Dentro de la enseñanza, los estudiantes aprendieron paso a paso a resolver problemas típicos de tareas bayesianas con la ayuda de una cierta representación. Los efectos de esta enseñanza fueron medidos por un test y un post test, con el fin de saber si ellos tenían éxito usando las representaciones aprendidas.

Los resultados mostraron que los estudiantes enseñados con árboles de frecuencia tuvieron un desempeño mucho mejor que estudiantes enseñados con

representaciones de probabilidad. Hay que notar que los desempeños del pre test fueron similares para los dos grupos, sin embargo las tasas de desempeño cambiaron en el post test siendo de 95% para los que usaban árboles y de 65% para los que usaban cálculos de probabilidades. En conclusión, los resultados experimentales apoyan la idea de que los formatos de frecuencia utilizan cálculos más sencillos que los requeridos por los formatos de probabilidad y esto facilita la resolución de los problemas de probabilidad condicional.

1.2.2.3. Lewis & Keren (1999). Haciendo referencia a quienes consideran que no es el formato sino el tipo de información que se da, lo que dificulta la solución de estos problemas, Lewis & Keren (1999) argumentan que existen dos factores diferentes en el contexto de las frecuencias: el uso de las frecuencias o de probabilidades y el uso de información conjunta. Para demostrar que el efecto del formato de frecuencias se debe al hecho de que la información se presenta conjuntamente, y no a través de condicionales, hacen un cruce de los enfoques de probabilidad con la forma en que se presenta la información, así:

	Intersecciones (I)	Condicionales (C)
Frecuencia (F)	Formato de frecuencia: FI	FC
Probabilidades (P)	PI	Formato de probabilidad: PC

Tabla 3. Combinaciones posibles para problemas de probabilidad condicional, según Lewis & Keren

Según estos autores, Gigerenzer & Hoffrage (1995) sólo consideraron dos casillas de la Tabla 3: FI y PC. Para la casilla FC los autores plantearon las condicionales frecuenciales que para el problema de la mamografía toman la siguiente forma:

10 de cada 1000 mujeres de 40 años que participan en un estudio tienen cáncer de seno. 800 de cada 1000 mujeres con cáncer de seno obtienen una mamografía positiva. 96 de cada 1000 mujeres sin cáncer de seno también obtienen una mamografía positiva. Una muestra representativa de mujeres obtuvo una mamografía positiva, ¿cuántas de estas mujeres se espera que tengan realmente cáncer de pecho? ____ de ____

Con esta nueva versión del problema, diseñan su estudio con dos problemas:

- El problema de la mamografía

Formato PC: La probabilidad de cáncer de seno es el 1% para una mujer de 40 que participa en un estudio de rutina. Si una mujer tiene cáncer de seno, la probabilidad de que obtenga una mamografía positiva es del 80%. Si una mujer no tiene cáncer de seno la probabilidad de que obtenga una mamografía positiva es 9.6%. Una mujer en un grupo de esta edad tiene una mamografía positiva en un estudio de rutina. ¿Cuál es la probabilidad de que actualmente tenga cáncer de seno?

Formato FI: 10 de cada 1000 mujeres de 40 años que han participado en un estudio de rutina tienen cáncer de pecho. 8 de cada 10 mujeres con cáncer de seno obtendrán una mamografía positiva. 95 de cada 990 mujeres sin cáncer de seno obtendrán una mamografía positiva. Se tiene una nueva muestra representativa de mujeres de 40 años quienes han recibido una mamografía positiva en un examen de rutina. ¿Cuántas de estas mujeres se esperaría que tuvieran actualmente cáncer de seno?

- El problema del taxi.

Un taxi se vio implicado en un accidente nocturno con choque y huida posterior. Hay dos compañías de taxis en la ciudad, la Verde y la Azul. El 85% de los taxis de la ciudad son Verdes y el 15% Azules. Un testigo identificó el taxi como Azul. El tribunal comprobó la fiabilidad del testigo bajo las mismas circunstancias que

había la noche del accidente y llegó a la conclusión de que el testigo identificaba correctamente cada uno de los colores en el 80% de las ocasiones y fallaba en el 20%. ¿Cuál es la probabilidad de que el taxi implicado en el accidente fuera en efecto Azul?

Y muestran que aún con frecuencias el razonamiento bayesiano es más difícil con información condicional que con información conjunta, y aseguran que el hecho de que una persona responda correctamente a un problema no necesariamente implica que use un razonamiento bayesiano.

Los problemas presentados por Lewis & Keren (1999) podrían verse como más difíciles respecto a Gigerenzer & Hoffrage (1995), ya que a pesar de tratar con frecuencias, éstas corresponden a un muestreo sistemático y no a un muestreo natural, como el de Gigerenzer & Hoffrage (1995).

En el Diagrama 4 se presenta la información dada por Lewis & Keren (1999). De entrada se ve que ésta es difícil de utilizar, ya que es como si se trabajara con probabilidades normales; inclusive sería más fácil tratar con los decimales que con los fraccionarios que se muestran. Opuesto a este diagrama se encuentra el Diagrama 5, que corresponde a la información dada por Gigerenzer & Hoffrage (1995) para el problema de la mamografía. Como ellos usan muestreo natural, se llega a un árbol en el que simplemente es necesario hacer operaciones como sumas con números naturales ($8+96$) y divisiones ($8/104$) para obtener el resultado pedido.

Luego Gigerenzer & Hoffrage (1995) no sólo presentan datos condicionales sino que operativamente son más sencillos de trabajar y son menos confusos para las personas.

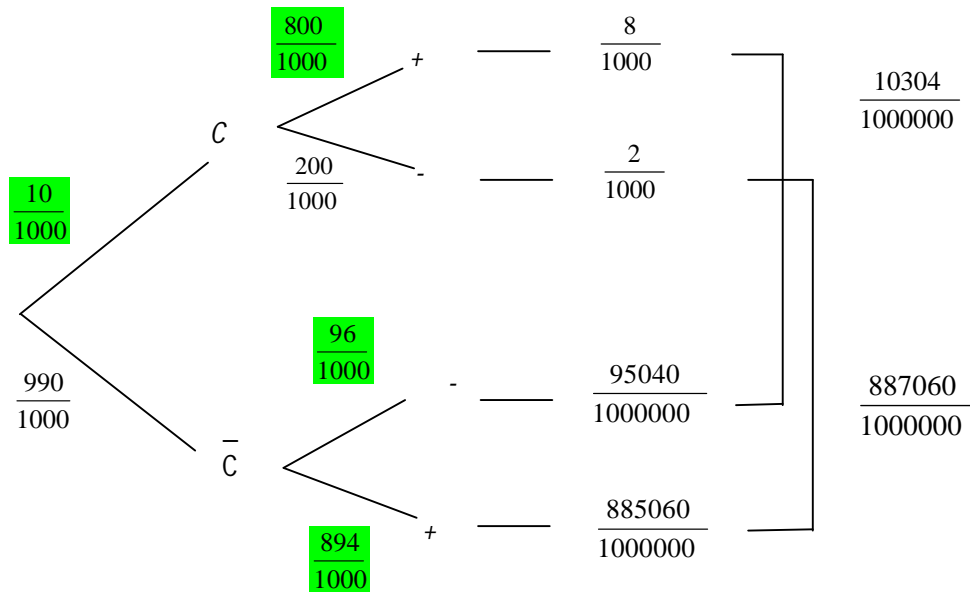


Diagrama 4. Árbol para las frecuencias de Lewis & Keren (1999) en el problema de la Mamografía.

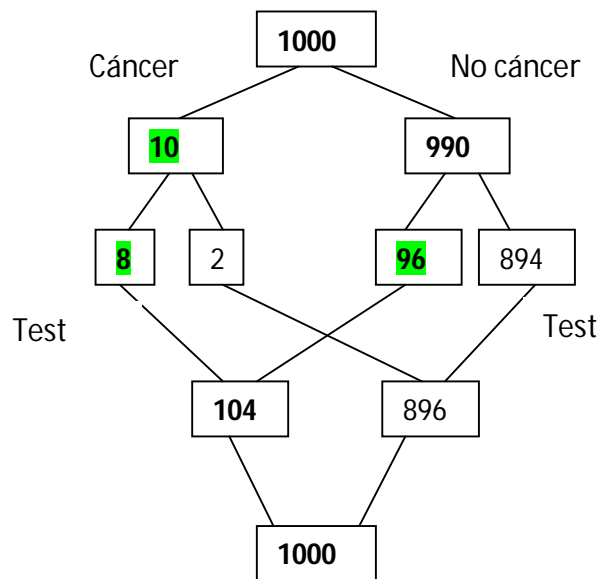


Diagrama 5. Árbol para las frecuencias de Gigerenzer & Hoffrage (1995) en el problema de la Mamografía.

1.2.2.4. Mellers & McGraw (1999). Estos autores muestran que el rendimiento se puede mejorar tanto con probabilidades como con frecuencias, dependiendo de lo extraños que sean los eventos y del tipo de información presentada. Cuando los eventos son raros, las probabilidades son más difíciles de entender que las frecuencias (esto, 5 de 1000 vs. 0.005). Además, cuando la información es presentada con intersecciones y marginales, se tiene una mejor idea de los conjuntos encajados.

Estos autores presentan seis versiones de un mismo problema, haciendo el cruce del tipo de información (estándar, menú corto), y de la forma de la información (probabilidades, muestreo sistemático y muestreo natural).

El muestreo sistemático² que no aparece en Gigerenzer & Hoffrage (1995) tiene la forma siguiente: La frecuencia de cáncer de seno es 1 de 100 mujeres con 40 años que participan en un estudio rutinario. En promedio, 80 de cada 100 mujeres con cáncer de seno obtienen mamografías positivas. En promedio, 9.6 de cada 100 mujeres sin cáncer de seno obtienen mamografías positiva. Hay una muestra representativa de mujeres de 40 años que obtienen mamografías positivas en un estudio de rutina. ¿Cuántas tienen realmente cáncer de seno?

El formato corto de frecuencias con muestreo sistemático tiene la forma: La frecuencia de cáncer de seno es 1 de 100 mujeres con 40 años que participan en un estudio rutinario. En promedio 0.8 de 100 mujeres tienen mamografía positiva y tienen cáncer de seno. En promedio 9.5 de cada 100 mujeres tienen mamografía positiva y no tienen cáncer de seno. Hay una muestra representativa de mujeres de 40 años que obtienen mamografías positivas en un estudio de rutina. ¿Cuántas tienen realmente cáncer de seno?

²En el muestreo sistemático la información se presenta sobre un total de 100 casos, 1000 o cualquier otra cantidad. Es decir, es el muestreo donde se fija la marginal en 100 y sobre este valor se consideran los datos.

Los resultados en el muestreo sistemático son ligeramente peores que en el muestreo natural, posiblemente porque los conjuntos encajados no son tan fáciles de visualizar en el primero de ellos.

Los autores realizaron dos estudios: en el primero se consideran eventos raros y 6 versiones del problema de la mamografía. Los resultados muestran grandes diferencias entre los formatos de probabilidad y los formatos de frecuencia. Dentro de los formatos de frecuencia son algo mejores los formatos de muestreo natural. Los autores reportan que con los formatos estándar se obtienen mejores respuestas que con los formatos cortos, tanto con probabilidades como con frecuencias, contrariamente a los resultados de Gigerenzer & Hoffrage (1995).

En el segundo estudio, los eventos considerados son más comunes y se obtienen resultados opuestos: el rendimiento depende del tipo de información más que de la forma. Ni las probabilidades ni las frecuencias producen consistentemente un mejor rendimiento. Las frecuencias en las tareas con información conjunta en muestro natural producen mejores respuestas (17%) que las probabilidades en el formato estándar (4%). No obstante, las probabilidades en formato corto producen un mejor porcentaje de respuestas acertadas (16%) que las frecuencias en formato estándar (8%).

En conclusión, dicen los autores, si tuviéramos que seleccionar una forma simple en la cual presentar información, recomendaríamos las frecuencias. Con eventos raros (5%), las frecuencias facilitan el razonamiento bayesiano. Si uno tuviera que seleccionar un solo tipo de información, recomendaríamos el menú corto o, más generalmente, cualquier tipo de información que permita a las personas visualizar conjuntos encajados.

Gigerenzer & Hoffrage (1995) responden a Lewis & Keren (1999) mostrándoles la imposibilidad de tener una información distinta a la conjunta en el formato de

frecuencias naturales, ya que precisamente en este formato siempre se obtienen las frecuencias de las intersecciones y no se requiere la regla del producto para calcularlas, como sí ocurre en el muestreo sistemático. Si el efecto del formato de frecuencias fuera debido fundamentalmente a que la información se presenta para las intersecciones, como Lewis & Keren (1999) afirman, entonces el formato de probabilidad con intersecciones conduciría a un rendimiento similar al obtenido con el formato de frecuencias. Pero no es el caso. Los resultados del estudio de 1995 muestran que con formato de probabilidades sólo se obtuvo 28% de respuestas bayesianas, en tanto que con el formato de frecuencias se alcanzó el 50%. En el mismo artículo, responden también a Meller & McGraw (1999) diciéndoles que efectivamente el efecto de las frecuencias se disminuye en el problema del taxi pero no en ninguno de los otros 14 problemas que ellos estudiaron. Una particularidad del problema del taxi, que puede explicar en alguna forma su dificultad y su poco éxito aún en formato de frecuencias, es dada por Yáñez (2005).

1.2.2.5. Evans & cols (2000). En su artículo: *Frequency versus probability formats in statistical Word problems*, Evans & cols (2000), dieron a conocer un estudio sobre formatos de presentación de problemas de probabilidad condicional en el que mostraron los resultados de tres experimentos y cómo éstos apoyaban la hipótesis de los autores en el sentido que no necesariamente las frecuencias naturales facilitan el proceso de solución de un problema de probabilidad condicional. Además, mostraron cómo el uso de frecuencias puede promover sesgos influenciados por la información.

Los experimentos se realizaron aplicando pruebas a quinientos estudiantes voluntarios de la universidad de Plymouth. En cada experimento se controlaron factores tales como: formato de la pregunta, formato de la tasa de falsos positivos, formato de la tasa base. Los problemas de la prueba eran de diagnóstico médico.

Un análisis de los resultados llevó a los investigadores a las siguientes conclusiones:

- Las versiones de frecuencia no son más fáciles que las versiones de probabilidad.
- Las respuestas erróneas disminuyen cuando en la redacción del problema se explica el significado de tasa de falsos positivos.
- El sesgo de la tasa base podría deberse a la congruencia entre el formato de la pregunta y el formato de la tasa base, es decir, si la pregunta y la tasa base se presentan en el mismo formato, la ocurrencia de este sesgo es mayor.
- La manera como los participantes responden está fuertemente influenciada por variaciones sutiles en la presentación de la tarea, la cual hace más o menos fácil la construcción de modelos mentales que ayudan a encontrar la solución.

1.2.3. Lonjedo y Huerta

En sus investigaciones sobre probabilidad condicional Lonjedo y Huerta, plantearon una clasificación de los problemas binarios de probabilidad condicional y exploraron aquellos factores que podrían influir en la solución de un problema. A continuación, se presentan las ideas generales de estas investigaciones y las conclusiones relevantes para esta investigación.

Lonjedo & Huerta (2004) y Lonjedo (2007), para facilitar el análisis didáctico que permita conocer mejor las dificultades de los estudiantes en cada tipo de problema, introducen a la clasificación dada por Yáñez (2001), la pregunta y la cantidad de marginales que se dan como información, como elementos

adicionales de clasificación del problema. De esta manera, definen dos nuevos criterios de clasificación:

Cuatro niveles: los mismos de Yáñez (2001).

Tres categorías:

$C_1 = 0$ marginales

$C_2 = 1$ marginal

$C_3 = 2$ marginales

Tres tipos:

Tipo 1: la pregunta es una probabilidad condicional

Tipo 2: la pregunta del problema es una probabilidad marginal

Tipo 3: la pregunta del problema es la probabilidad de una intersección.

La combinación de niveles, tipo y categorías produce los diferentes problemas en la nueva clasificación, que se sintetiza en la Tabla 4, donde N es nivel y N_i hace referencia al número de condicionales. En principio cada nivel queda dividido en 9 clases, dependiendo de la categoría y del tipo. Pero, como la categoría depende del nivel en el que se encuentra y el tipo está ligado a si el problema sea de probabilidad condicional, en la Tabla 4 las casillas en blanco indican la imposibilidad de considerar problemas pertenecientes a las clases correspondientes a esas casillas.

	N_1			N_2			N_3			N_4		
C_1	C_1T_1			C_1T_1	C_1T_2	C_1T_3	C_1T_1	C_1T_2	C_1T_3	C_1T_1	C_1T_2	C_1T_3
C_2	C_2T_1			C_2T_1	C_2T_2	C_2T_3	C_2T_1	C_2T_2	C_2T_3			
C_3	C_3T_1			C_3T_1		C_3T_3						

Tabla 4. Tipos de problemas de probabilidad condicional en la clasificación de Lonjedo & Huerta (2004)

Como complemento a esta clasificación, hicieron estudios sobre las actuaciones de los estudiantes al resolver un problema de probabilidad condicional (Lonjedo & Huerta, 2003, 2004, 2006, 2007a, 2007b y Lonjedo, 2007). Lo primero que hicieron los investigadores fue recoger, de los enunciados de los problemas, expresiones del lenguaje natural no simbólico que inducen a interpretaciones no deseadas en los datos y en las preguntas, con el fin de eliminar la ambigüedad de los textos de los problemas de probabilidad condicional y ayudar al éxito de los participantes en el problema.

En sus investigaciones (Lonjedo & Huerta, 2006, 2007a, 2007b), los autores concluyen que la expresión *de los que...*, mejora la comprensión del dato sobre la probabilidad condicional respecto de las otras formulaciones.

También investigaron sobre la influencia del formato de presentación del problema en la resolución del mismo y encontraron que los problemas presentados en términos de frecuencias fueron más fáciles que los de probabilidad y porcentajes, aun para estudiantes en niveles educativos básicos; estos resultados concuerdan con los de Gigerenzer & Hoffrage (1995), pero agregan que en los problemas en términos de frecuencias cuando se da un dato que se refiere a probabilidades condicionales es mejor darlo como porcentaje ya que así es posible diferenciarlo de las frecuencias absolutas, o naturales, mientras se realiza su lectura.

Los problemas utilizados fueron de nivel 2 (N2), es decir, con una condicional según la clasificación de Lonjedo & Huerta, 2006, 2007a, 2007b (los mismos que en la clasificación de Yáñez, 2001). La elección de estos problemas se debió a que este nivel es el que menos se encuentra en los textos de educación secundaria en España y, al no poder justificar esta ausencia, quisieron verificar si estos problemas pueden ser resueltos por estudiantes de diferentes niveles educativos. En estos problemas, además, se tuvieron en cuenta diferentes variables de tarea, como la naturaleza de los datos, las estructuras condicionales

del lenguaje natural no simbólico, tanto en el dato como en la pregunta del problema, así como el número de relaciones aditivas y multiplicativas necesarias para su resolución.

En estos estudios participaron 256 estudiantes de diferentes niveles educativos, desde 4° ESO a 2° bachiller, así como estudiantes universitarios de la facultad de matemáticas.

Uno de los aspectos tratados en las investigaciones fue la influencia del formato en la resolución de los problemas de probabilidad condicional

Algunas de las conclusiones que permitieron los resultados fueron:

- El presentar el dato condicional como porcentaje, cuando todos los demás datos se dan en frecuencias naturales, contribuye positivamente a disminuir el porcentaje de estudiantes que confunden la probabilidad condicional con la conjunta, es decir, este es un buen camino para empezar a eliminar esta confusión.
- El porcentaje de estudiantes que resuelve correctamente el problema es mayor para el caso de los porcentajes que en el de probabilidades. Ellos contradicen de alguna manera lo dicho por Gigerenzer & Hoffrage (1995) que no encontraron diferencias entre ellos.
- Se sugiere que el proceso de la enseñanza de la probabilidad condicional comience con problemas cuya estructura y formato de los datos sean las frecuencias en contextos variados, como un paso previo para abordar problemas en porcentajes y finalizar con problemas en términos de probabilidades. La razón es que así se introduce a los estudiantes a la probabilidad condicional mediante razones y proporciones, siguiendo con el uso de los porcentajes con cierto sentido probabilístico y finalizando con el uso de las probabilidades, pues este proceso contribuye con uno de los

objetivos de la enseñanza de la probabilidad condicional, de que quien resuelve sea capaz de escoger entre las diferentes herramientas utilizadas para resolver un problema de probabilidad condicional.

- Los datos presentados adquieren significado para los estudiantes cuando están expresados en porcentajes y frecuencias absolutas, ya que los estudiantes pueden producir nuevas cantidades relevantes para la solución del problema que facilitan el proceso de resolución.

COMENTARIOS FINALES

El resumen que se ha realizado de algunos estudios matemáticos, psicológicos y didácticos más relevantes relacionados con el razonamiento condicional, permite llegar a las siguientes conclusiones:

1. El teorema de Bayes como se propone en (3), no adquiere mayor interés para el formato corto propuesto por Gigerenzer & Hoffrage (1995), debido a que en éste sólo se requieren dos probabilidades conjuntas para encontrar una probabilidad *a posteriori*. Por el contrario, adquiere interés en problemas de formato estándar, ya que así se evidencia que el cálculo de las probabilidades *a posteriori* en este formato se realiza en función de las probabilidades *a priori* y de las verosimilitudes, donde las probabilidades $P(A_i)$ son probabilidades *a priori*, las probabilidades $P(A_i | B)$ son las probabilidades *a posteriori* y las probabilidades $P(B | A_i)$ son las verosimilitudes.
2. Uno de los factores que ha llevado a dividir las posiciones respecto al formato de frecuencias es la concepción que cada corriente tiene sobre ellas. Por una lado, están aquellos que hablan de frecuencias en el sentido

de muestreo natural y por el otro, quienes consideran las frecuencias como un muestreo sistemático. En este trabajo se acoge la idea de muestreo natural, de acuerdo con Gigerenzer & Hoffrage (1995)

3. Los problemas de probabilidad condicional no son sencillos de resolver, pues el proceso de solución implica diversas dificultades tal como lo presenta Díaz & de la Fuente (2006). Además, traducir del lenguaje natural al simbólico no es una tarea fácil.
4. Entre todos los errores que se pueden cometer al resolver problemas de probabilidad condicional, se destaca uno que es casi universal y es la confusión de la probabilidad condicional con la conjunta.
5. Las representaciones utilizadas para resolver un problema de probabilidad condicional pueden facilitar, o por el contrario causar dificultades, a la hora de dar solución al problema. Existe una estrecha relación entre la representación en diagrama de árbol con la forma secuencial en que se procesa la información.
6. Existe una facilidad a la hora de resolver un problema de probabilidad condicional proporcionada por las frecuencias naturales que como lo afirma Gigerenzer & Hoffrage (1995) están relacionadas con la forma secuencial en la que se adquiere la información.
7. Las expresiones del lenguaje utilizadas para expresar la condicionalidad no deben ser ambiguas, pues de entrada se estaría obstaculizando el proceso de razonamiento llevando a las personas a confundir la probabilidad condicional con la conjunta y a cometer otro tipo de errores.

8. Responder correctamente a un problema de probabilidad condicional no significa que las personas estén usando un razonamiento correcto durante todo el proceso de resolución.
9. El tipo de información que se proporciona en un problema de probabilidad condicional no se ha estudiado a fondo como factor influyente en el éxito de la solución a un problema de probabilidad condicional.

CAPÍTULO II

METODOLOGÍA

Este capítulo responde a inquietudes que surgen a todo lector interesado en esta investigación y que son cruciales en el momento de juzgar la validez de la misma. Dichas inquietudes están relacionadas con la forma como ésta se realizó, es decir, el tipo de investigación, la descripción de la muestra, las fases de la investigación y el análisis de los procesos de resolución de los problemas estudiados. Los dos primeros aspectos se presentan en la primera parte de este capítulo, los otros se describen detalladamente en la segunda parte.

2.1. TIPO DE INVESTIGACIÓN Y DESCRIPCIÓN DE LA MUESTRA.

El principal objetivo de esta investigación fue estudiar el efecto que tienen los diferentes formatos y el tipo de información que se presenta en un problema binario de probabilidad condicional³ sobre las respuestas y estrategias que las personas utilizan para resolverlo. Para alcanzarlo, se diseñó una investigación que proporcionara nuevo conocimiento en el campo de la enseñanza de la probabilidad condicional y que pueda ayudar tanto a profesores como estudiantes a comprender y superar las dificultades que se presentan al resolver un problema de probabilidad condicional. A continuación se muestran aspectos generales de esta investigación.

2.1.1. Tipo de investigación. Dada la complejidad que implica el estudio de las dificultades que se presentan al resolver un problema binario de probabilidad condicional y el efecto del formato y el tipo de información dada, se seleccionaron dos métodos de investigación: uno descriptivo con el que se pretendía identificar y analizar las dificultades que presentaron los estudiantes al resolver los problemas aquí planteados. El otro, de carácter cuantitativo, con el fin de cuantificar el efecto del formato y el tipo de información en el éxito de la solución a un problema de

³ Se llama problema de probabilidad condicional a un problema de probabilidad en el que, o bien en los datos, o bien en la pregunta del problema, se incluye una probabilidad condicional.

probabilidad condicional por medio de pruebas para comparar proporciones y gráficos de interacción.

2.1.2. Población y muestra. La población objeto de estudio son los estudiantes universitarios que han tomado un curso de probabilidad y estadística. La muestra estuvo conformada por 432 estudiantes activos de la Universidad Industrial de Santander (UIS) del primero y segundo semestre académico del 2009 de las carreras de Licenciatura en Matemáticas, Economía, Ingeniería de Sistemas, Ingeniería Industrial, Ingeniería Eléctrica, Ingeniería Civil e Ingeniería Electrónica. La Tabla 5, presenta la distribución de los estudiantes según cada carrera.

Carrera	Curso	Estudiantes	Total
Licenciatura en Matemáticas	Estadística I	13	56
	Estadística II	20	
	Didáctica de la probabilidad y la estadística	23	
Economía	Estadística I	21	21
Ingeniería Industrial	Estadística I	50	90
	Estadística III	40	
Ingeniería de Sistemas	Estadística II	68	68
Ingeniería Electrónica	Probabilidad y Estadística	62	62
Ingeniería Eléctrica	Probabilidad y Estadística	35	35
Ingeniería Civil	Estadística aplicada a la ingeniería	100	100
			432

Tabla 5. Distribución de los estudiantes de acuerdo a la cantidad de cursos tomados de probabilidad y estadística que presentaron las pruebas de estudio

2.2. FASES DE LA INVESTIGACIÓN Y ANÁLISIS DE LOS PROCESOS DE RESOLUCIÓN. El objetivo general y los antecedentes permitieron plantear las siguientes hipótesis iniciales de investigación:

- Los problemas que se presentan en formato de frecuencias naturales producen mayor porcentaje de respuestas correctas.
- Los problemas que presentan probabilidades condicionales en su enunciado producen menos éxito que aquellos en que no se presentan sino que se preguntan.
- El lenguaje en que se expresa la condicionalidad es otro factor que influye en el éxito de solución.
- El éxito en la solución de un problema de probabilidad condicional también está influenciado por el tipo de representación que se elija.

Para comprobar estas hipótesis, la investigación se desarrolló en tres fases que se describen a continuación.

2.2.1. Primera fase. Estudio piloto. La primera fase contó con un estudio piloto de propósito múltiple: generar una idea a priori sobre la dificultad de los problemas binarios de probabilidad condicional y evaluar la redacción de los problemas con el ánimo de mejorarla en las pruebas de estudio y de esta manera obtener resultados más significativos en la investigación.

En esta fase participaron 50 estudiantes de la carrera de Licenciatura en Matemáticas de la UIS que se encontraban tomando un curso de estadística. 32 estudiantes se enfrentaron a la prueba piloto 1 [Anexo A] que presenta 5 problemas pertenecientes a los casos 1, 2, 3, 4 y 5 según la clasificación de Yáñez

(2001); los 18 restantes presentaron la prueba piloto 2 [Anexo B] que contaba con 4 problemas de los casos 6, 7, 8 y 9. El tiempo de presentación de las pruebas fue de una hora.

Del análisis de los procedimientos mostrados por los estudiantes en las pruebas piloto se decidió que los problemas a incluir en las pruebas de estudio serían de nivel 0 y 1, específicamente los casos 2 y 5 debido a que en estas pruebas presentaron mayores índices de respuesta y también porque se quería evaluar un caso por cada nivel. Los casos de niveles 2 y 3 no se incluyeron en las pruebas de estudio porque la cantidad de estudiantes que los respondieron correctamente no fue significativa.

2.2.2. Segunda fase. Estudio sobre el efecto del formato y el tipo de información. En esta fase se elaboraron 4 problemas [Anexo C], todos del mismo contexto y con una misma pregunta pero diferente tipo de información y formato de presentación, de la siguiente manera:

Problema 1: presentado en formato de frecuencias y caso 5.

Problema 2: presentado en formato de probabilidades y caso 5.

Problema 3: presentado en formato de frecuencias y caso 2.

Problema 4: presentado en formato de probabilidades y caso 2.

Participaron 240 estudiantes (60 por ítem), a los que se les dio para que resolvieran, sin previo aviso y de manera aleatoria, uno de los problemas. Los estudiantes contaron con 10 minutos para resolverlo. Además para que la variable profesor no tuviera efecto, en cada curso se distribuyeron los cuatro problemas.

Para analizar los datos obtenidos al aplicar los problemas individuales, primero se realizó una comparación entre las proporciones de respuestas correctas para el tipo de información y el formato de presentación con el fin de observar y concluir

cuál de estos factores causaba un mayor efecto sobre las respuestas de los estudiantes.

Para este análisis fue indispensable definir los siguientes elementos:

VARIABLES:

- DE RESPUESTA: proporción de que respondieron correctamente al problema.
- DE ESTUDIO: tipo de problema y formato de presentación del problema.
- A CONTROLAR: Todos los estudiantes que presentan la prueba han visto un curso de probabilidad en un tiempo no superior a un año.

HIPÓTESIS ESTADÍSTICAS:

Ho: No hay diferencia en las proporciones de respuestas correctas entre los formatos de presentación.

Ha: la proporción de respuestas correctas para las frecuencias naturales es mayor que la de probabilidades.

Ho: No hay diferencia en las proporciones de respuestas correctas entre el tipo de información.

Ha: la proporción de respuestas correctas para el caso 2 es mayor que para el caso 5.

Una vez realizada la comparación de proporciones y probadas las hipótesis, se realizó un gráfico de interacción (p. 70), cuyo objetivo era mostrar el efecto en la respuesta al presentar un problema en un formato y con un tipo de información específica, es decir, el efecto de la interacción de los dos factores sobre las respuestas de los estudiantes.

Los resultados de esta etapa se encuentran en la primera sesión del capítulo de resultados y análisis.

2.2.3. Tercera fase. Estudio sobre las dificultades. La última fase consistió en la aplicación de dos pruebas de estudio [Anexos D y E] con las que se identificaron y analizaron las concepciones, intuiciones, dificultades y errores que los estudiantes presentaron al resolver dichos problemas.

La necesidad de la aplicación de dos pruebas, surgió porque en la primera los problemas 5 y 6 presentaron dos dificultades:

1. Eran el mismo problema sólo que en diferente formato de presentación, esto hizo que varios estudiantes interpretaran de igual forma los dos problemas.
2. La suma de los datos que se daban como información era 100% o 1 y esto contribuía para que los estudiantes confundieran con mayor facilidad la probabilidad condicional con la conjunta.

Por estos motivos se elaboró una segunda prueba en donde se cambiaron los ítems 5 y 6.

Estas pruebas se aplicaron, sin previo aviso, a 192 estudiantes de la UIS pertenecientes a las carreras ya mencionadas, 77 de los cuales presentaron la prueba de estudio 1 y 115 la prueba de estudio 2. Estas pruebas fueron aplicadas en una hora de clase y su presentación, por iniciativa del profesor del curso, fue de carácter obligatorio.

Las pruebas de estudio 1 y 2 contenían seis problemas binarios de probabilidad condicional, tres de ellos pertenecientes al caso 2 y los otros al caso 5. Los tres problemas de cada caso eran diferentes en cuanto al contexto y cada uno se presentó en un formato diferente: en porcentajes, en probabilidades y en

frecuencias naturales. Además cada problema sólo tenía una pregunta para que el tiempo concedido para la solución de la prueba fuera suficiente.

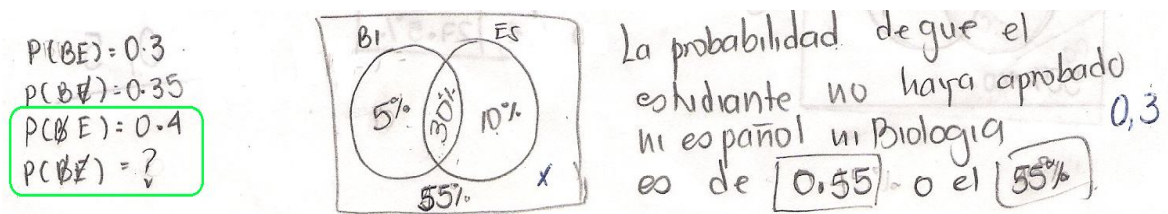
Con el fin de analizar los resultados de estas pruebas, fue necesario organizar las respuestas dadas por los estudiantes según las características que mostraron en el proceso de resolución de cada problema. La siguiente lista resume las dificultades y fortalezas que presentaron y una explicación de cada una de ellas.

- *Respuesta correcta*: tanto las representaciones como la respuesta y los procedimientos son correctos.
- *Identificación correcta de los datos*: sólo identifica correctamente los datos dados en la parte informativa del problema pero no responde correctamente a la pregunta.
- *No responde*: no muestra ningún proceso de resolución.
- *No identifica datos o identificación incorrecta*: realiza una identificación errónea de los datos que se presentan ya sea en la parte informativa del problema o en la pregunta, o simplemente no logra identificarlos.
- *Representación incorrecta*: hace representaciones incorrectas de los datos, especialmente en lo que se refiere a los diagramas de Venn, tablas de doble entrada y diagramas de árbol.
- *Confunde probabilidad condicional y conjunta*: interpreta $P(A|B)$ como $P(A \cap B)$ o viceversa.
- *Confunde probabilidad marginal con condicional*: interpreta $P(A|B)$ como $P(A)$ o $P(B)$.
- *Confunde una probabilidad condicional con su inversa*: interpreta $P(A|B)$ como $P(B|A)$ o toma $P(A|B) = P(B|A)$.
- *Confunde un suceso con su complementario*: interpreta $P(B)$ como $P(\bar{B})$.

- *Error en la partición del espacio muestral:* el estudiante considera que las probabilidades que se dan como información corresponden a todo el espacio muestral o que entre ellas son complementarias.
- *Fallo con las operaciones:* errores en las operaciones aritméticas (suma, resta, multiplicación o división).
- *Dificultad en la fórmula de Bayes:* se presentan expresiones matemáticas que no corresponden a la fórmula de Bayes.
- *Error en conceptos y propiedades de la probabilidad:* el estudiante asimila que una probabilidad puede ser mayor que uno o escribe expresiones erróneas de las fórmulas básicas (4), (5) y (6) presentadas en el capítulo de antecedentes.

Además de estas dificultades, también se analizó el efecto del lenguaje utilizado para expresar la condicionalidad, el contexto de los problemas y el tipo de representación utilizada como medio para ilustrar mejor la información, ya que estos aspectos también influyen a la hora de resolver exitosamente un problema de probabilidad condicional.

Una vez revisadas las pruebas y definidas estas dificultades, se clasificó a cada estudiante en alguna de ellas y se tomó como dato el número de estudiantes que presentó cada dificultad. Teniendo en cuenta que un mismo estudiante podía cometer más de un error, para poder clasificarlo dentro de una sola dificultad se escogió aquella que lo llevó a la respuesta incorrecta o a desencadenar otros errores. Por ejemplo, la Evidencia 2.1, muestra dos errores en la actuación del estudiante: primero confunde la probabilidad condicional con la conjunta y segundo hace una representación incorrecta de los datos. Este estudiante fue clasificado dentro de los que confunden la probabilidad condicional con la conjunta ya que fue este error el que lo llevó a interpretar mal los datos y la pregunta, además fue el causante de que elaborara una mal representación.



Evidencia 2.1.

Los porcentajes de estudiantes que se clasificaron dentro de cada dificultad se encuentran en la Tabla 7, dentro del capítulo de resultados y análisis (p. 72). Los datos allí presentados, se analizaron de manera cualitativa, pero adicionalmente en el ítem 1 se realizó una prueba de comparación de proporciones que se necesitó para establecer si la diferencia entre el porcentaje de estudiantes que respondieron correctamente a este ítem en las dos pruebas, fue significativa. Estos análisis dieron origen a las conclusiones y comentarios realizados en relación a cada una de las dificultades.

Los resultados de esta tercera fase se encuentran en la segunda sesión del capítulo de resultados y análisis.

CAPÍTULO III

RESULTADOS Y ANÁLISIS

Este capítulo contiene los resultados más importantes de esta investigación. El capítulo está dividido en dos secciones. En la primera se estudia el efecto de los formatos de presentación y el tipo información de los problemas sobre las respuestas de los estudiantes; este análisis se realiza en base a los resultados de los problemas individuales. En la segunda sección se hace un análisis de las dificultades que presentan los estudiantes al resolver los problemas binarios de probabilidad condicional de las pruebas de estudio 1 y 2. A medida que se presentan los resultados de la investigación, se realizan las observaciones y comentarios correspondientes a cada uno de ellos.

3.1. EFECTO DEL FORMATO Y TIPO DE INFORMACIÓN

Para estudiar el efecto de los formatos y el tipo de información sobre las respuestas de los estudiantes, se decidió hacer pruebas de proporciones con los datos obtenidos al aplicar las pruebas individuales con el fin de realizar análisis objetivos de los datos que condujeran a deducciones válidas respecto al problema de estudio.

En la Tabla 6 se muestran las proporciones de respuestas correctas obtenidas para cada uno de los problemas individuales clasificados de acuerdo al caso al cual corresponde y al tipo de formato.

FORMATO CASO	FRECUENCIAS NATURALES	PROBABILIDADES	TOTAL
CASO 2	12/60	14/60	26/120
CASO 5	5/60	0/60	5/120
TOTAL	17/120	14/120	31/240

Tabla 6. Proporción de respuestas correctas, según tipo de información y formato.

Una primera observación que se puede hacer a los resultados de la Tabla 6, es la baja proporción de respuestas correctas, únicamente 31 estudiantes resolvieron el

problema correctamente (13%). Otra observación importante es que el tipo de formato no tuvo un efecto significativo sobre el éxito al resolver el problema en ninguno de los dos casos. Lo contrario sucedió con el tipo de problema, es así que se aprecian diferencias entre las proporciones de respuestas correctas a los problemas de caso 2 y caso 5, contando con mucho más éxitos los problemas del caso 2. En pocas palabras y de manera a priori se puede afirmar que el tipo de información tiene algún efecto sobre las respuestas de los estudiantes.

La prueba que se presenta a continuación es para mostrar que sí existen diferencias significativas entre las proporciones de respuestas correctas para diferentes tipos de información.

Las hipótesis de la prueba son:

$$H_0: \hat{p}_1 = \hat{p}_2$$

$$H_a: \hat{p}_1 > \hat{p}_2$$

\hat{p}_1 proporción de respuestas correctas para caso 2

\hat{p}_2 proporción de respuestas correctas para caso 5,

Utilizando el software FATHOM para hacer la prueba se obtuvieron los siguientes resultados:

In **caso 2** 26 out of 120 , or 0,216667, are **Category**

In **caso 5** 5 out of 120 , or 0,0416667, are **Category**

Alternative hypothesis: The population proportion for **Category** in **caso 2 is greater than** that for **Category** in **caso 5**

The test statistic **z**, is **4,042**.

If it were true that the two population proportions were equal (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value of **z this great or greater** would be **< 0,0001**

Note: This probability was computed using the normal approximation.

Según muestran los resultados con un valor $p < 0,0001$ y $\alpha = 0,05$ se rechaza la hipótesis nula, por lo que se concluye que las diferencias entre las proporciones

para los dos casos son significativas, lo que permite decir que el caso 2 produce una proporción de respuestas correctas mayor que la que produce el caso 5. Además, como el caso 2 pertenece al nivel 0 y el caso 5 es de nivel 1, de acuerdo con la clasificación de Yáñez (2001), se puede añadir que los problemas de nivel 0 tienen mayor proporción de respuestas correctas que los del nivel 1, resultado que está en la dirección planteada por el mismo Yáñez (2001) cuando afirma que la dificultad de un problema de probabilidad condicional es mayor cuanto mayor sea el número de probabilidades condicionales que contiene en su parte informativa. Resta para próximas investigaciones probar si esto se cumple para los demás niveles y con los casos del mismo nivel.

A continuación se presentan los resultados de una prueba para comparar las proporciones de respuestas correctas según formato de presentación:

La hipótesis para la prueba de la comparación de proporciones es:

$$H_0: \hat{p}_1 = \hat{p}_2$$

$$H_a: \hat{p}_1 > \hat{p}_2$$

\hat{p}_1 proporción de respuestas correctas para el formato de frecuencias naturales

\hat{p}_2 proporción de respuestas correctas para el formato de probabilidades,

In **frecuencias naturales** 17 out of 120 , or 0,141667, are **Category**

In **probabilidades** 14 out of 120 , or 0,116667, are **Category**

Alternative hypothesis: The population proportion for **Category** in **frecuencias naturales** is greater than that for **Category** in **probabilidades**

The test statistic: **z**, is **0,5774**.

If it were true that the two population proportions were equal (the null hypothesis), and the sampling process were performed repeatedly, the probability of getting a value of **z** **this great or greater** would be **0,28**.

Note: This probability was computed using the normal approximation.

Con los resultados proporcionados por el software, se tiene un valor p de 0.28 y que con un $\alpha = 0,05$, no existe evidencia suficiente para rechazar la hipótesis nula, es decir, las diferencias entre las proporciones de respuestas correctas entre los formatos no son estadísticamente significativas. Por lo que se concluye que para los problemas escogidos para esta investigación, no se puede decir que las

frecuencias naturales producen una proporción de respuestas correctas significativamente mayor que la que producen las probabilidades.

Para ver la interacción que se da entre el formato y tipo de información, se elaboró el Gráfico 1, en el cual se puede ver que no hay una combinación dentro de los problemas que sea la mejor, si lo que se quiere es tener mayor porcentaje de respuestas problemas de tipo 2 en cualquiera de los formatos es aconsejable, sin embargo los problemas de tipo 5 son sensibles al cambio de formato; haciendo una prueba para comparar las dos proporciones según el tipo de formato y con las siguientes hipótesis:

$$H_0: \hat{p}_1 = \hat{p}_2$$

$$H_a: \hat{p}_1 > \hat{p}_2$$

\hat{p}_1 proporción de respuestas correctas para el formato de frecuencias naturales y tipo 5

\hat{p}_2 proporción de respuestas correctas para el formato de probabilidades y tipo 5.

Se obtuvo un valor p de 0.0113 y con un $\alpha=0,05$, luego se rechaza la hipótesis nula, es decir, que los problemas de tipo 5 presentados en formatos de frecuencias producen más respuestas correctas que los de el mismo tipo pero en formato de probabilidades.

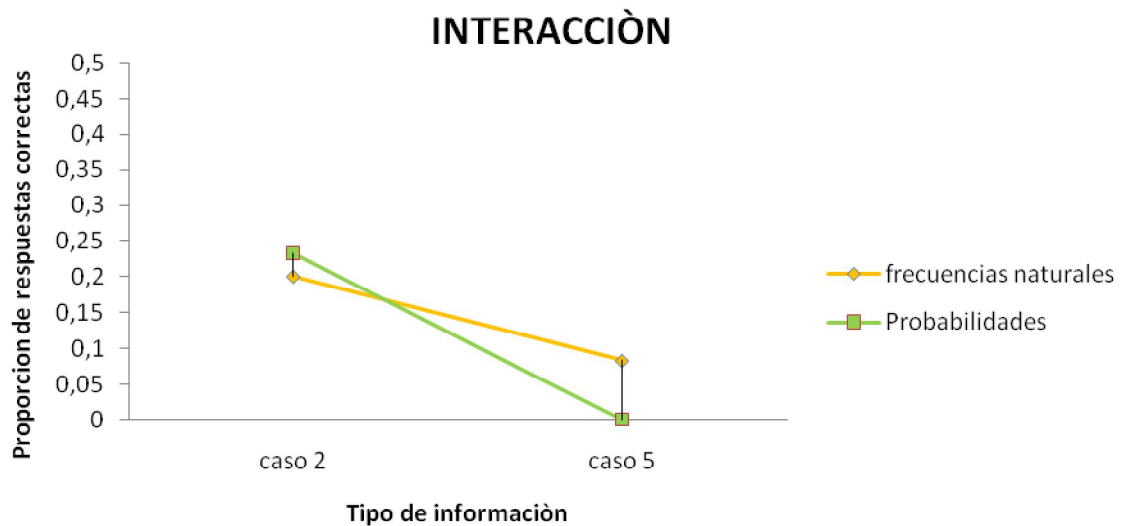


Gráfico 1. Interacción entre formato y tipo de información.

3.2. ANÁLISIS DE LAS DIFICULTADES EN LA SOLUCIÓN DE LOS PROBLEMAS BINARIOS DE PROBABILIDAD CONDICIONAL.

Hacer un estudio de problemas de probabilidad condicional necesariamente conduce a un análisis de los procedimientos llevados a cabo por las personas para obtener su solución. Algunos autores en sus estudios (Assumpta, Díaz & de la Fuente, 2006; Falk, 1986; Kanheman & Tversky, 1982; Huerta & Lonjedo, 2003a) han encontrado errores que se cometen al resolver este tipo de problemas, tales como la falacia de la condicional transpuesta y la confusión de la probabilidad condicional y la conjunta, y viceversa. Pero estos errores no son necesariamente debidos a una incapacidad de las personas para interpretar correctamente la información condicional, existen otros factores como el lenguaje en que se presenta el problema y aún el contexto del mismo que pueden influir en la solución, tal como afirma Yáñez (2001):

“Las diversas razones esgrimidas para explicar estas dificultades se pueden clasificar en dos grandes grupos: la primera son las razones psicológicas...la segunda, son las razones del lenguaje y hacen referencia a la gran cantidad de formas y contextos con que se pueden presentar los problemas de probabilidad condicional...” (p.356).

Es lógico pensar que los errores más comunes que se comenten al resolver un problema binario de probabilidad condicional disminuyen en la medida en que la persona recibe instrucción al respecto, es decir, se está de acuerdo con Martignon & Wassner (2002) en que tales dificultades se pueden superar, siempre y cuando no existan otros factores que dificulten el entendimiento.

En este sentido, esta sección está dirigida a mostrar un análisis de las dificultades presentadas por los estudiantes al enfrentarse a problemas de probabilidad condicional, teniendo en cuenta las estrategias de resolución de los problemas en las pruebas de estudio 1 y 2. A continuación se describen en detalle cada una de las dificultades encontradas ejemplificándolas con las respuestas de algunos estudiantes y se realizan las observaciones pertinentes.

La Tabla 7, presenta los porcentajes de estudiantes clasificados en cada una de las dificultades.

	Ítems comunes prueba 1 y 2			Prueba 1			Prueba 2		
	ITEM 2	ITEM 3	ITEM 4	ITEM 1	ITEM 5	ITEM 6	ITEM 1	ITEM 5	ITEM 6
Respuesta correcta	37,7	11,27	80,31	55,8	16,9	18	35,7	7,83	15,7
No respondió	3,74	14,32	3,04	0	2,6	10	0	24,3	22,6
Identificación correcta de los datos	2,41	9,55	0,87	3,9	5,19	3,9	1,74	6,09	6,09
No identifica datos o identificación incorrecta	7,64	5,00	3,25	3,9	5,19	1,3	2,61	10,4	8,7
Representaciones incorrectas	1,5	1,52	1,96	0	2,6	3,9	0	4,35	1,74
Confunde probabilidad condicional y conjunta	22,74	23,66	0	23,4	53,2	48	52,2	6,96	27,8
Confunde probabilidad marginal con condicional	0,65	8,89	0	3,9	6,49	3,9	0,87	2,61	4,35
Confunde una probabilidad condicional con su inversa	0,44	1,08	0	5,19	0	1,3	5,22	11,3	0
Fallo con las operaciones	1,08	0,65	2,39	1,3	0	0	1,74	0	1,74
Dificultad con fórmula de Bayes	1,77	2,82	0	0	0	0	0	1,74	0,87
Concepto y propiedades de la probabilidad condicional	19,75	18,66	6,74	2,6	6,49	7,8	0	21,7	10,4

Tabla 7. Distribución de porcentajes por ítem para las diferentes dificultades.

Al observar la Tabla 7 hay un dato que llama la atención en el ítem 4, donde se observa un 80.31% de respuestas correctas, de entrada genera una incoherencia con las conclusiones obtenidas en la sección anterior, por lo que a continuación se da una explicación a este hecho.

El tener un 80.31% de respuestas correctas en un ítem de nivel 1 de por sí es muy alto, además al comparar éste con el 56% que corresponde al máximo porcentaje de respuestas correctas para ítems de nivel 0, parece contradecir de manera directa la conclusión a la que se llegó con las pruebas de proporciones, pero no es así, ya que lo primero que se debe tener en cuenta es que el formato de

presentación del ítem 4, es de frecuencias naturales y en este formato las cantidades que se refieren a información condicional son las mismas que hacen referencia a una intersección. Por ejemplo 20 mujeres no casadas (intersección) son las mismas 20 mujeres entre los no casados (condicional). Por lo que se tendría que un problema de caso 5, nivel 1, se puede transformar en un problema de nivel 0, caso 1, si se toma el dato que hace referencia a una condicional y se interpreta como una intersección, así se tendrían tres intersecciones. En general cualquier problema que presente un dato condicional (niveles 1, 2 y 3 según la clasificación de Yáñez, 2001) en este formato (frecuencias naturales), puede convertirse en un problema de nivel 0.

Por otra parte el hecho de que ninguna persona haya confundido la condicional con la conjunta en este ítem se justifica porque realmente al presentarse el problema en formato de frecuencias naturales ésta deja de ser una confusión. Aquí aparecen algunas ventajas de las frecuencias naturales sobre las probabilidades que se expondrán en el capítulo de conclusiones.

Ítem 4: Se sortea un viaje a Singapur entre los 120 mejores clientes de una agencia de automóviles. De ellos 45 son mujeres casadas, 35 son hombres casados y además 20 de los que no están casados son mujeres.
¿Cuál es la probabilidad de que el afortunado sea hombre?

3.2.1. Dificultades con el lenguaje utilizado para expresar la condicionalidad.

El éxito o el fracaso en la resolución de problemas de probabilidad condicional no depende únicamente de la estructura y los conocimientos de probabilidad. Lonjedo (2007), afirma que la estructura semántica de un problema binario de probabilidad condicional influye en la resolución del mismo, de hecho presentan una serie de construcciones gramaticales que favorecen la interpretación de los datos y de las preguntas relacionadas con la probabilidad condicional, mostrando

que el conector condicional utilizado por excelencia (SI) tiene mejores resultados que otros; sin embargo, añaden:

“Construir todos los problemas de probabilidad condicional utilizando únicamente este conector, sería empobrecedor y concluyen que posiblemente existen otros conectores (entre lo que..., de los que...) que inducen al estudiante a la interpretación deseada de la condicionalidad, siempre y cuando la oración condicional esté bien construida y no presente ambigüedades”. (p. 197)

Este estudio contó con unos ítems que permitieron no sólo evidenciar el lenguaje como un factor influyente sino también confirmar lo planteado por Lonjedo & Huerta, 2006, 2007a, 2007b.

Las pruebas de estudio 1 y 2 fueron aplicadas a dos grupos de estudiantes diferentes ($N_1= 77$ y $N_2= 115$). El ítem 1 fue similar en las dos pruebas lo único que cambió fue la forma de la pregunta. En el caso de la prueba 1 se utilizó el conector *entre los...*, que facilitó la comprensión de los estudiantes con un 56% de respuestas correctas, sin embargo para la prueba 2 no se utilizó ningún conector que especificara la condicionalidad, aspecto que seguramente generó confusión entre los estudiantes e hizo que el porcentaje de respuestas correctas fuera de 36%. Haciendo una prueba para comparar estas proporciones se encontró una diferencia significativa ($p=0,0029$), lo que es un claro indicativo de que el lenguaje utilizado para expresar la condicionalidad influyó en la solución del problema. En este caso, el cambio en la estructura de la pregunta indujo a los estudiantes a confundir la probabilidad condicional con la conjunta, confusión que se estudiará más adelante. Los enunciados para el ítem 1 fueron:

(Prueba de estudio 1). Un centro escolar con 1000 alumnos entre niños y niñas tiene 282 estudiantes que usan gafas, 147 niñas que las usan y 368 niñas que no las usan.

Entre los niños ¿qué porcentaje usa gafas?

(Prueba de estudio 2). Un centro escolar con 1000 alumnos entre niños y niñas tiene 282 estudiantes que usan gafas, 147 niñas que las usan y 368 niñas que no las usan.

¿Qué porcentaje de niños usa gafas?

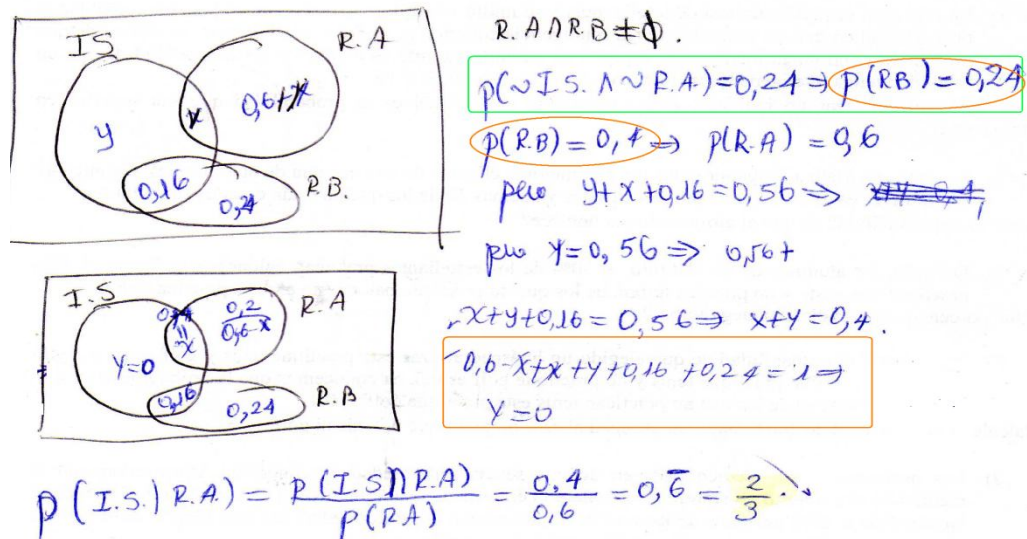
3.2.2. Dificultades con el contexto del problema. El contexto de un problema es un aspecto fundamental para la solución del mismo, por ejemplo, si a un estudiante se le da un problema cuyo contexto parece contradictorio con su intuición, éste aunque conozca las propiedades de la probabilidad seguramente no logrará ajustarlas al problema. Tal fue el caso del ítem 3, en el que muchos estudiantes manifestaban un error en el problema pues para ellos era imposible que una persona con inteligencia superior tuviera un rendimiento bajo, “es lógico” decían algunos que si tiene inteligencia superior debe tener un rendimiento alto y no pueden existir personas con inteligencia baja y rendimiento alto, además que esa repetición de palabras es una retahíla que confunde.

Ítem 3: En un grupo se realizó un test de inteligencia y se midió su rendimiento académico y se encontró que la probabilidad de que un individuo tenga inteligencia superior es de 0.56, además la probabilidad de que un individuo no tenga inteligencia superior ni rendimiento alto es de 0.24 y la probabilidad de que un individuo tenga inteligencia superior y rendimiento bajo es de 0.16.

Si se selecciona al azar un individuo con rendimiento alto, ¿Cuál es la probabilidad que sea superior en inteligencia?

Un estudiante que manifestó esta dificultad fue Andrés, quien durante la prueba todo el tiempo insistió en lo ilógico del problema y como se ve en su actuación [Evidencia 3.1.] él intenta hacer un diagrama para representar la información y al no tener éxito con éste decide resolver el problema desde el álgebra; lo primero

que se puede apreciar en su procedimiento es, que tener rendimiento bajo e inteligencia inferior es sinónimo de tener rendimiento bajo (0.24). Además no logra conectar su intuición con sus resultados, pues tiene dos valores de probabilidad para rendimiento bajo (0.24 y 0.4) uno que se corresponde con la intuición y otro que tiene que ver con los valores de su diagrama.



Evidencia 3.1.

Este problema de contexto explica el por qué del bajo porcentaje de respuesta correcta (11%) comparado con el 46% y el 38% (Tabla 7) para los ítems 1 y 2 respectivamente, a pesar de que estos pertenecen al tipo 2 en la clasificación de Yáñez (2001).

El ítem 5 de la prueba de estudio 2, también tuvo dificultades ligadas al contexto del problema, puesto que los eventos que en éste se trataban tenían que ver con situaciones cotidianas como color del cabello y color de ojos. En este problema las personas ligaban su experiencia a lo que consideraban podría ser la respuesta, por ejemplo en nuestro medio es más común ver una persona con ojos castaños y cabellos castaño, que una persona con cabello castaño y ojos azules. Por esta razón para las personas fue difícil asimilar los datos que se daban.

(Ítem 5, Prueba de estudio 2): En cierta ciudad el 50% de las personas no tienen ojos castaños ni cabello castaño, mientras que el 15% de ellas, tienen cabello castaño y ojos castaños. Se sabe además que el 37.5% de los que tienen cabellos castaño tienen ojos castaños.

¿Qué porcentaje de las personas en esta ciudad tienen ojos castaños?

3.2.3. Problemas con la definición de probabilidad condicional y sus propiedades. Resolver problemas de probabilidad condicional requiere además de comprender el problema, conocer, entender y aplicar exitosamente las relaciones básicas de la teoría de las probabilidades. Los estudiantes saben que existen estas relaciones, pero muchas veces caen en el error de ignorar su utilidad y creen que cuando se les da una fórmula sólo deben aprender a insertarle datos sin comprender y analizar los resultados a la luz de la teoría y del contexto del problema. Este es otro de los errores que lleva a los estudiantes a fracasar en la resolución de los problemas de probabilidad condicional.

En todos los problemas utilizados en esta investigación, los estudiantes dejaron traslucir dificultades en la comprensión de las fórmulas básicas de las probabilidades, los porcentajes de estudiantes identificados fueron para el caso de los ítems 2, 3 y 5 del 20%, 19% y 14% respectivamente [Tabla 7].

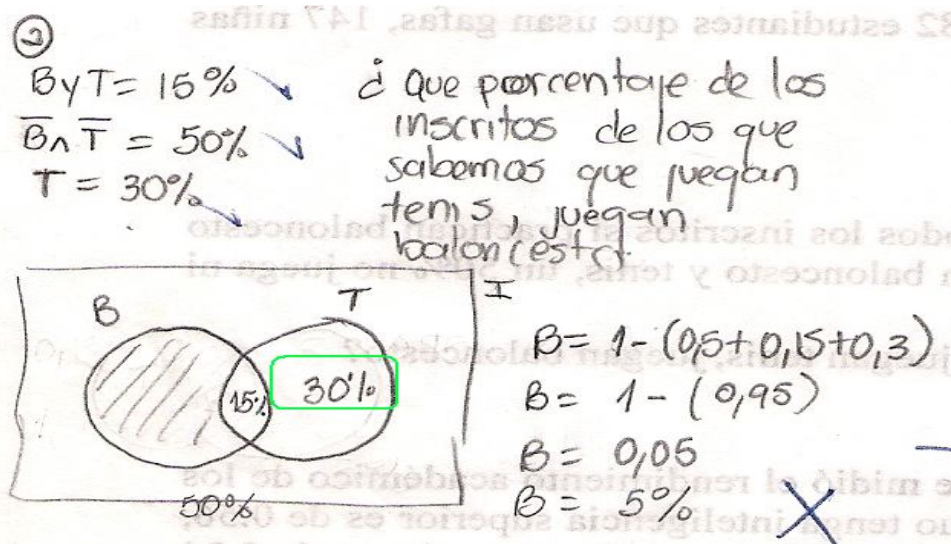
Dentro de esta dificultad se incluyeron errores tales como:

- ❖ *Confundir la probabilidad marginal con la conjunta.* Este error lo cometió el 20% de los estudiantes, porcentaje considerable si se tiene en cuenta que dentro de los artículos estudiados en este trabajo no se encontraron referencias al respecto, posiblemente porque en estos estudios el error apareció pero con poca frecuencia.

Analizando las actuaciones de los estudiantes con este error, se observó que el causante de la confusión estuvo en el ítem 2, en el cual se daba la probabilidad de jugar tenis (marginal) y era interpretada por los estudiantes como la probabilidad de jugar tenis y no jugar baloncesto (conjunta), como lo refleja la Evidencia 3.2.

Ítem 2: Para organizar un campeonato deportivo hemos preguntado a todos los inscritos si practican baloncesto y/o tenis. Hemos encontrado que un 15% de los inscritos juegan baloncesto y tenis, un 50% no juega ni baloncesto ni tenis y un 30% juega tenis.

¿Qué porcentaje de los inscritos al campeonato, de los que sabemos que juegan tenis, juegan baloncesto?



Evidencia 3.2

- ❖ *Olvido de las fórmulas básicas de la probabilidad condicional, definidas en el capítulo de antecedentes.* En su intento por aplicar las fórmulas aprendidas sobre la probabilidad condicional, los estudiantes crean nuevas relaciones que en alguna forma reemplazan a las fórmulas correctas. Algunos ejemplos son:

$$P(A \cap B) = P(A) - P(B)$$

$$P(A) = P(A|B) + P(A|\bar{B}) \quad \text{En reemplazo de } P(A) = P(A \cap B) + P(A \cap \bar{B})$$

Una curiosidad que se puede ver en este reemplazo es que como

$$P(A|B) + P(A|\bar{B}) = \frac{P(A \cap B)}{P(B)} + \frac{P(A \cap \bar{B})}{P(\bar{B})}.$$

Y como es común que las personas al sumar fracciones cometan el error de sumar numerador con numerador y denominador con denominador, entonces como $P(B) + P(\bar{B}) = 1$ se tiene que $P(A|B) + P(A|\bar{B}) = P(A \cap B) + P(A \cap \bar{B})$ y esta es la relación que finalmente utilizan los estudiantes.

- ❖ *Generalización de la fórmula utilizada para la intersección de eventos independientes como válida para dos eventos cualesquiera.* Entre todos los estudiantes el 5.2% asumieron que la probabilidad de ocurrencia conjunta de dos eventos era igual al producto de las probabilidades marginales de dichos eventos como lo muestra la Evidencia 3.3 olvidando que esto sólo se puede aplicar cuando los eventos son independientes.

Handwritten work showing a student's incorrect calculation of joint probability as the product of marginal probabilities:

$$P(\text{INT. SUP} \wedge \text{REND. BAJO}) = 0,16$$

$$P(\text{INT. SUP}) * P(\text{REND. BAJO}) = 0,16$$

$$0,56 * P(\text{REND. BAJO}) = 0,16$$

$$P(\text{REND. BAJO}) = \frac{0,16}{0,56} = 0,285$$

Evidencia 3.3

3.2.4. Confusión de la probabilidad conjunta con la condicional y viceversa.

Los estudiantes confunden las probabilidades condicionales con las probabilidades conjuntas, y las conjuntas con condicionales (Ojeda, 1995; Díaz & de la Fuente, 2006; Díaz, Ortiz & Serrano, 2007; Pollatsek et all, 1987; Yáñez,

2003; Lonjedo, 2007). A continuación se muestra cómo en este estudio se dieron los dos casos. En casi todos los ítems los porcentajes de estudiantes que cometen alguno de estos errores fue considerable, superior al 20% en todos los ítems, excepto el ítem 4 y 5 de la prueba 2 que fueron del 0 y 7%, como se muestra en la Tabla 7.

En la Tabla 7, se puede apreciar que los porcentajes de personas que confunden la probabilidad condicional con la conjunta y viceversa varían de una prueba a otra en algunos ítems como es el caso del 1, 5 y 6. Para el ítem 1, esta diferencia de porcentajes de una prueba a otra se debió principalmente al lenguaje utilizado en la pregunta para expresar la condicionalidad, como ya se dijo.

Para el ítem 5 el origen de esta diferencia es otro, ya que son dos problemas diferentes en su contexto en las dos pruebas y esto afecta las respuestas de los estudiantes. Otra observación que se debe hacer es que para la prueba 1 únicamente el 2.6% de estudiantes no respondieron este ítem mientras que en la prueba 2 este porcentaje fue de 24.3%, esta diferencia entre los porcentajes de no respuesta influye en la cantidad de estudiantes que confunden la probabilidad condicional con la conjunta, ya que si hay más estudiantes que contestan el ítem hay mayor probabilidad de que cometan errores. Una tercera observación al respecto es que de todos los errores cometidos por los estudiantes en el ítem 5, el de confundir la probabilidad conjunta con la condicional fue el mayor para la prueba 1, pero no para la prueba 2 cuyo error más frecuente tuvo que ver con las propiedades y la definición de probabilidad condicional (21.7%) [Ver Tabla 7].

(Ítem 5, Prueba de estudio 1): De todos los alumnos de un instituto, el 30% de los estudiantes practican baloncesto y fútbol, el 30% practica baloncesto y no practica fútbol, de los que no practican baloncesto el 40% practica fútbol.
¿Qué porcentaje de estudiantes juega fútbol?

El caso del ítem 6 es similar al del 5, ya que en la prueba 1 solamente el 10% no responde el ítem mientras que en la 2 este porcentaje fue de 27.8%. Además hay que notar que para la prueba 1 este ítem era semejante al 5, sólo que uno estaba en formato de porcentajes y el otro en formato de probabilidades; los estudiantes notaron esta semejanza y en muchos casos si resolvían bien el primero también resolvían el segundo, pero si fallaban en el primero también fallaban en el segundo y así pues al equivocarse en la interpretación del dato en el problema 5 también lo hacían en el 6, de aquí que los porcentajes de estudiantes que confunden la probabilidad condicional con la conjunta sean similares para estos ítems en la prueba 1 (53% y 48%) [Tabla 7]

(ítem 6, Prueba de estudio 1): En un hotel la probabilidad de que elegido un huésped al azar éste practique tenis y golf es de 0.3 y la probabilidad de que practique tenis y no practique golf es 0.3. Si conocemos que la probabilidad de que elegido un huésped de los que no practican tenis este practique golf es 0.4. Calcule la probabilidad de que tomado un huésped al azar no practique ni tenis ni golf.

(ítem 6, Prueba de estudio 2): En un salón de clases la probabilidad de que al seleccionar aleatoriamente un estudiante éste haya aprobado biología y español es de 0.3 y la probabilidad de que haya aprobado biología y no haya aprobado español es de 0.35. Se sabe además que si se selecciona aleatoriamente un estudiante entre los que perdieron biología la probabilidad de que este haya aprobado español es de 0.4. Calcule la probabilidad de que elegido un estudiante al azar en el salón de clases éste no haya aprobado biología ni español.

3.2.4.1. Confusión entre probabilidad conjunta y condicional. Esta dificultad hace referencia al hecho tomar un dato que representa una intersección y confundirlo o interpretarlo como una condicional, lo mismo puede suceder cuando

se pregunta por una intersección y se responde con una condicional. En este estudio hubo dos ítems: 3 y 6 de la prueba 2, en los que se manifestó esta confusión. La actuación de Carlos [Evidencia 3.4] ilustra claramente esta confusión ya que él asumió las probabilidades conjuntas que se presentaban en la información como condicionales. Aunque para el caso del ítem 3 (p. 75) este no fue el único inconveniente con que se encontraron los estudiantes, ya que como se analizó antes el contexto de este problema fue confuso.

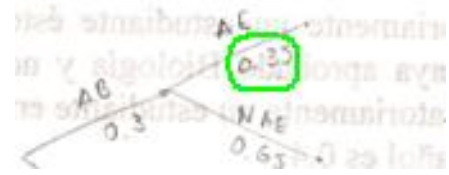
Otra manifestación de esta confusión se presentó al construir los diagramas de árbol; por ejemplo en la solución dada por Andrea [Evidencia 3.5] si sólo se mira el diagrama es claro que ella ubica las intersecciones donde deben estar las condicionales, sin embargo no se puede decir si el problema está en que la estudiante definitivamente confunde la probabilidad conjunta con la condicional o no sabe en realidad construir la representación. Este es otro de los problemas que tienen los estudiantes, pues ellos conocen la forma de la representación pero no discriminan los datos que deben usar para construirla correctamente y poder usarla aplicando las reglas que el profesor enseñó.

En este punto cabe resaltar que en su mayoría los problemas utilizados para la enseñanza de la probabilidad condicional son del tipo 7 (Yáñez, 2001) y como se analizó en los antecedentes (p. 36) los datos que se dan se ajustan muy bien a los diagramas de árbol. Además, desde el punto de vista pedagógico enseñar siempre un mismo tipo de problema de una u otra forma prepara al estudiante para aprender un algoritmo de solución que él considera válido para resolver cualquier problema, sin importar el tipo de información que se dé [ver anexo F].

$$P(\text{no intel} / \text{rendimiento alto}) = 0,24$$

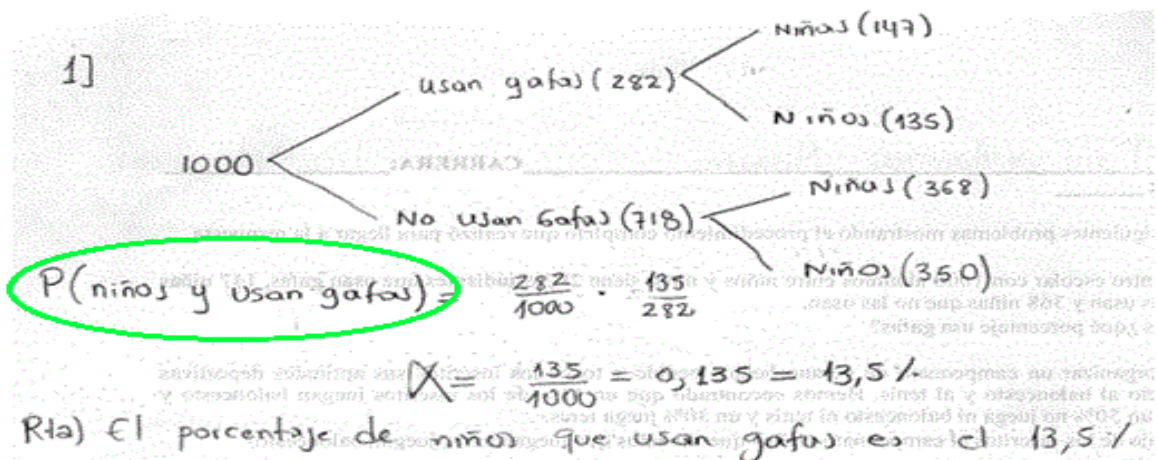
$$P(\text{no intel} / \text{rendimiento bajo}) = 0,16$$

Evidencia 3.4.



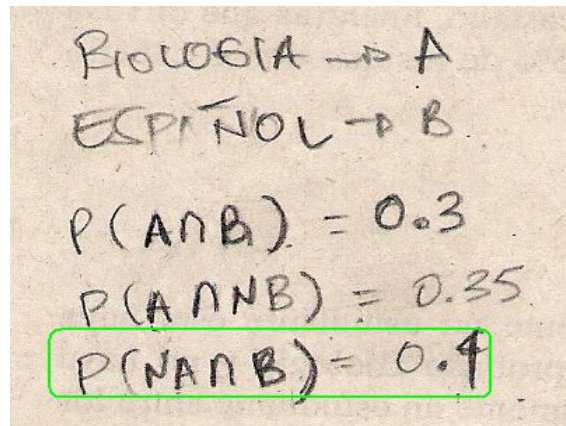
Evidencia 3.5.

3.2.4.2. Confusión entre probabilidad condicional y conjunta. Esta dificultad tiene que ver con el hecho de confundir información que representa probabilidades condicionales con intersecciones. Diversos estudios han presentado evidencias respecto a esta confusión (Lonjedo, 2007; Díaz & de la Fuente, 2006; Díaz, Ortiz & Serrano, 2007), en éste también se muestran; el ítem 1 de la prueba 2 presentó esta dificultad al momento de responder a la pregunta, pues como se dijo antes no había un conector que definiera la condicionalidad. Un ejemplo de esto es la solución presentada por Raúl [Evidencia 3.6], quien elaboró un diagrama de árbol correcto pero falló al interpretar la probabilidad que se pedía en la pregunta ya que da como respuesta la probabilidad de niños que usan gafas cuando se le había preguntado por la probabilidad de usar gafas dado que es niño.



Evidencia 3.6.

El ítem 6 de la prueba 2, como ya se dijo también presentó un alto porcentaje de personas que confundieron la probabilidad condicional con la conjunta, la Evidencia 3.7 presenta de manera clara cómo el estudiante confunde la probabilidad de aprobar español dado que perdió biología (condicional) con la probabilidad de aprobar biología y perder español (conjunta).


$$\begin{aligned} \text{BIOLOGIA} &\rightarrow A \\ \text{ESPAÑOL} &\rightarrow B \\ P(A \cap B) &= 0.3 \\ P(A | B) &= 0.35 \\ P(B | A) &= 0.4 \end{aligned}$$

Evidencia 3.7.

3.2.5. Sobre las representaciones. De acuerdo al procedimiento sugerido por Díaz, Ortiz & Serrano (2007) para resolver un problema de probabilidad condicional, construir una representación adecuada es uno de los primeros y fundamentales pasos, dado que una buena representación permite tener una idea más clara del problema, pues al elaborarla el estudiante muestra de manera explícita aquello que piensa, al tiempo que debe identificar los sucesos que intervienen en el problema. El asunto está en que la escogencia de una determinada representación está ligada al hecho de que ésta se ajuste a la información que el problema proporciona, puesto que a la hora de resolver un problema de probabilidad condicional las personas pueden utilizar tablas de doble entrada, diagramas de árbol, diagramas de Venn o la representación algebraica y todas se ajustan a unos problemas mejor que a otros. En este sentido el trabajo de Yáñez (2001) es suficientemente explicativo ya que caracteriza los problemas de probabilidad condicional a la luz de la representación utilizada. En consecuencia,

es válido pensar que una buena representación influye en la correcta solución del problema. Por estas razones, es importante hacer un análisis sobre el tipo de representación utilizada por los estudiantes para resolver cada ítem de acuerdo con el tipo de información que se daba y cómo se relaciona con los porcentajes de respuestas correctas y las dificultades presentadas. La Tabla 8, presenta los porcentajes de personas que utilizaron cada representación en cada ítem.

Tipo de representación	ítem 1	ítem 2	ítem 3	ítem 4	prueba 1		prueba 2	
					ítem 5	ítem 6	ítem 5	ítem 6
No responde	0,0	4,2	14,1	3,6	2,6	11,7	25,2	25,2
Tabla de doble entrada	4,7	3,6	9,9	10,9	13,0	11,7	2,6	2,6
Álgebra o aritmética	82,3	49,5	47,9	64,1	23,4	31,2	52,2	49,6
Diagrama de Venn	3,6	37,0	5,2	2,1	57,1	42,9	10,4	7,8
Diagrama de Árbol	9,4	5,7	22,9	19,3	3,9	2,6	9,6	14,8

Tabla 8. Distribución de porcentajes de representación por ítem, para las pruebas de estudio 1 y 2.

En términos generales, lo primero que se puede apreciar en la Tabla 8 es que en todos los ítems la representación más utilizada fue el álgebra o la aritmética, excepto en los ítems 5 y 6 de la prueba de estudio 1.

En ítems como el 1 y el 4, las personas en su mayoría prefieren utilizar la aritmética para resolver el problema, pues por la manera en que se dan los datos estos permiten hacer sumas y divisiones de manera sencilla, se podría considerar de acuerdo con Lonjedo & Huerta (2004), que estos serían más bien problemas de asignación de probabilidades, en los que no es necesario hacer uso de las propiedades de la probabilidad para resolverlo.

Del análisis hecho en los antecedentes (p. 38), sobre el tipo de representación más apropiada para los problemas de tipo 2, se llegó a que una tabla de doble entrada o un diagrama de Venn, permitían una mejor ubicación de los datos. Para el ítem 1, sin embargo por el formato de presentación del problema (frecuencias

naturales), la aritmética constituye una estrategia no sólo válida sino efectiva para resolverlo, ya que de todos los estudiantes que lo respondieron correctamente (44%) un 83% (70 estudiantes) de ellos usaron la aritmética, como se observa en el Gráfico 2.

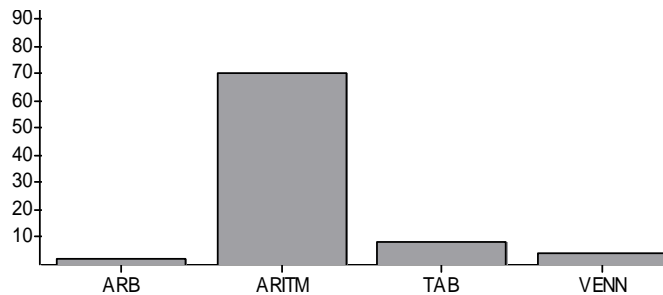


Gráfico 2. Diagrama de barras de la representación utilizada para el ítem 1, de los estudiantes que respondieron correctamente.

En el ítem 2, también del tipo 2, un diagrama de Venn o una tabla de doble entrada, serían las representaciones más recomendadas, sin embargo el 50% (ver Tabla 8 de los estudiantes, escogieron la aritmética y como se aprecia en el Gráfico 3, esta representación no tuvo mucho éxito ya que únicamente 35% de los estudiantes que la usaron respondió correctamente el ítem, para el caso del diagrama de Venn (barras verdes) del 37% que lo usó el 40%, resolvió el ítem con éxito. Por lo que los resultados apoyan a los diagramas de Venn como una mejor representación de los datos.

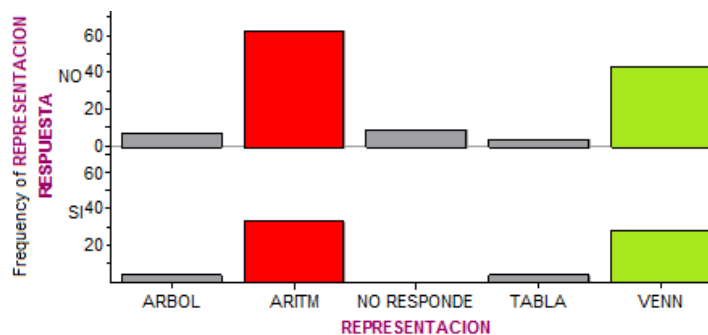


Gráfico 3. Distribución de las representaciones utilizadas en el ítem 2 (Si es respuesta correcta. No es respuesta incorrecta).

Para el ítem 3, la recomendación respecto a la representación es la misma que para los ítems 1 y 2, pero como la información se daba en formato de probabilidad los estudiantes usaron en mayor medida los diagramas de árbol, comparado con el uso que de éstos se hizo en los ítems 1 y 2. El Gráfico 3., muestra que del 23% de estudiantes que usaron los diagramas de árbol el 98% (barras verdes) falló al responder el ítem. Es curioso el hecho de que los estudiantes que confundieron la probabilidad condicional con la conjunta fue de 23.66%, casi el mismo de los que usaron diagramas de árbol, lo que sugiere que en efecto fue el haber utilizado los diagramas de árbol lo que ayudó a la aparición de la confusión.

Otra representación muy utilizada aquí fue la aritmética, pero esta no produjo mucho éxito, como lo muestra el Gráfico 4, son más los que respondieron mal (43%, barra roja) el ítem usando esta representación que los que lo hicieron bien (0.5%).

En este ítem el problema de contexto que ya se mencionó, unido a una inadecuada representación produjo un alto porcentaje de respuestas incorrectas.

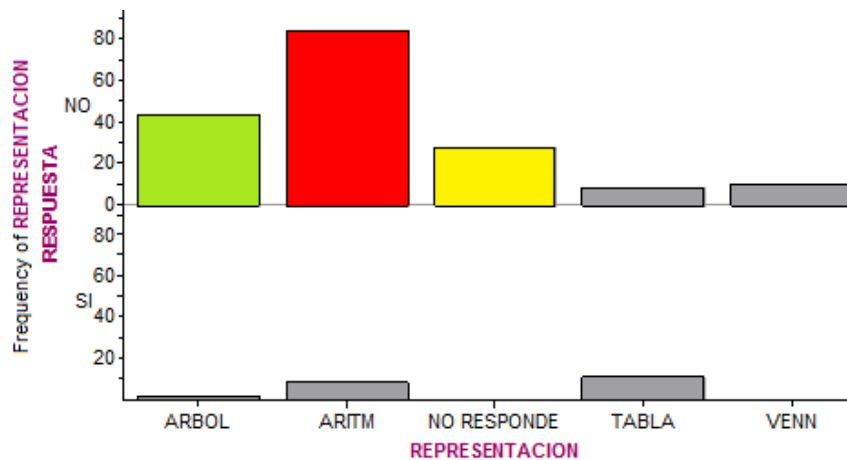


Gráfico 4. Distribución de las representaciones utilizadas en el ítem 3 (Si es respuesta correcta. No es respuesta incorrecta).

En el ítem 4, no se determinó una representación como la más adecuada, pero al poseer datos que hacen referencia a probabilidades condicionales, un diagrama de árbol es una mejor opción que las tablas, ya que éstas no permiten ver las

condicionales. Pero sin duda nuevamente por ser el formato frecuencias naturales la aritmética funciona bien, como lo indican los resultados, del 64% de personas que hizo uso de la aritmética un 44% respondió correctamente. Otra representación que se usó en este ítem fue el diagrama de árbol que como lo muestra el Gráfico 5, de todos los que usaron diagrama de árbol (19%), el 78% lo resolvió de manera correcta.

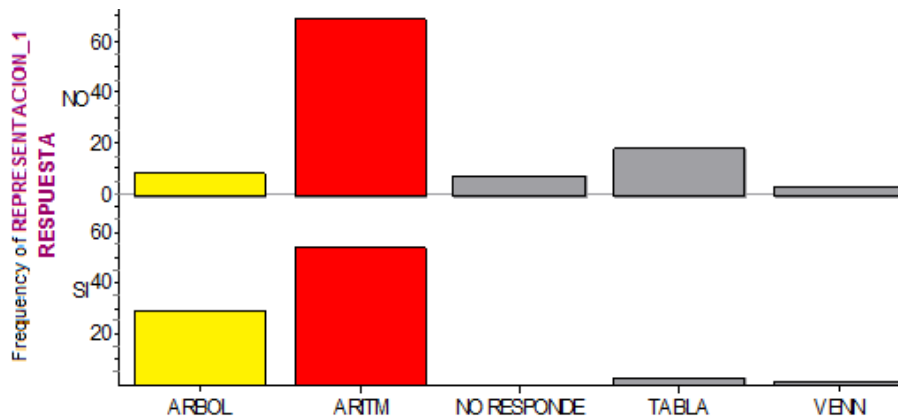


Gráfico 5. Distribución de las representaciones utilizadas en el ítem 4 (Si es respuesta correcta. No es respuesta incorrecta).

En los ítems 5 y 6 de la prueba 1, predominaron los diagramas de Venn sobre la aritmética, como se ve en la Tabla 8. Para estos ítems la representación recomendada es un diagrama de árbol, porque en éste el dato condicional se puede ubicar de manera visible. Sin embargo como lo muestran los Gráficos 6 y 3.7, el porcentaje de estudiantes que hizo un diagrama de Venn y llegó a una respuesta correcta fue de 10% para el ítem 5 y 14% para el ítem 6, contrario a esto se tiene que el porcentaje de respuestas incorrectas fue de 47% para el ítem 5 y 29% para el ítem 6. Lo que es un claro indicativo de que este tipo de representación lleva a cometer errores cuando se tiene problemas que involucran probabilidades condicionales.

Cuando se presentó la confusión de la probabilidad condicional con la conjunta, estos ítems presentaron porcentajes muy altos de confusión. Esto unido al hecho

de que las personas hubiesen preferido un diagrama de Venn, pudo ser un factor que ayudó a aumentar dicha confusión. De aquí la importancia en la escogencia de la representación, pues una mala representación lleva en muchos casos a una respuesta errada del problema.

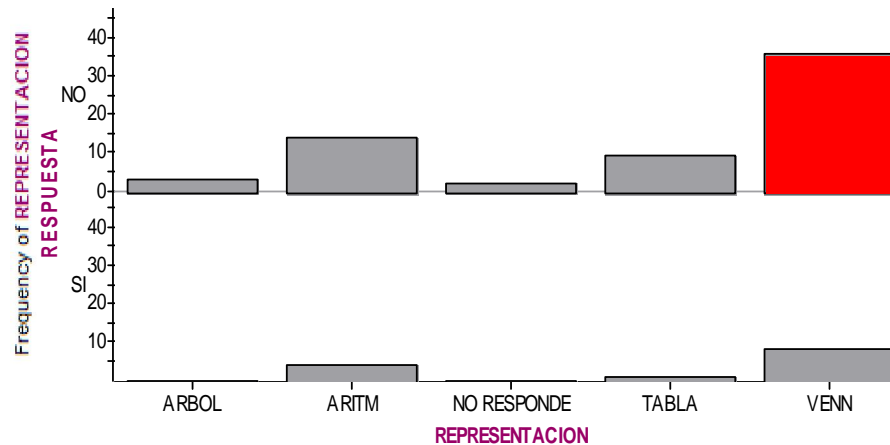


Gráfico 6. Distribución de las representaciones utilizadas en el ítem 5, prueba 1 (Si es respuesta correcta. No es respuesta incorrecta).

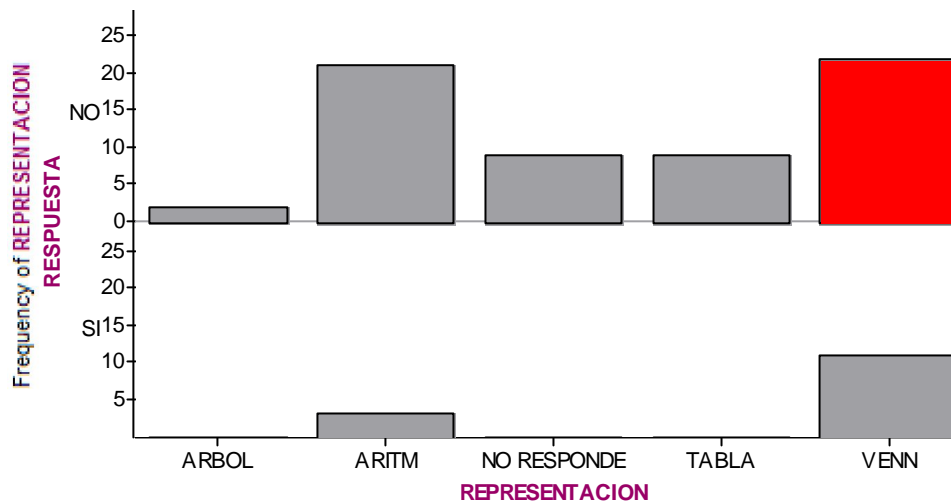


Gráfico 7. Distribución de las representaciones utilizadas en el ítem 6, prueba 1 (Si es respuesta correcta. No es respuesta incorrecta).

En los ítems 5 y 6 de la prueba de estudio 2, la representación sugerida es un diagrama de árbol, que permite ubicar la probabilidad condicional dada en la

información. Para el ítem 5 sólo 11 estudiantes utilizaron los diagramas de árbol [ver Gráfico 8, barras verdes] de los cuales 1 respondió correctamente, de nuevo los estudiantes prefirieron la aritmética [ver Gráfico 9, barras rojas) 60 de ellos la usaron, de los cuales únicamente 6 respondieron correctamente.

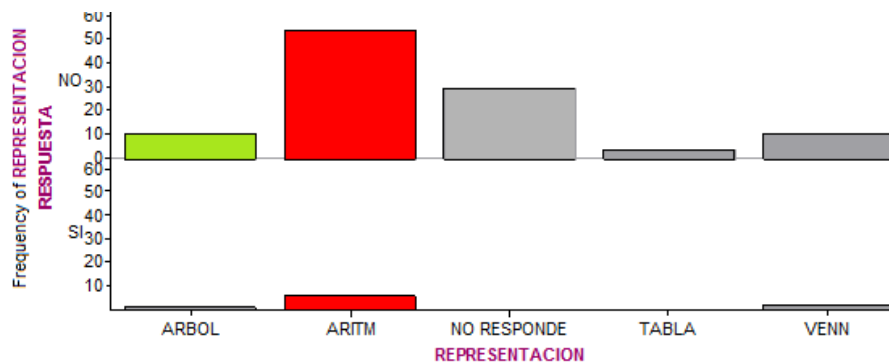


Gráfico 8. Distribución de las representaciones utilizadas en el ítem 5, prueba 2 (Si es respuesta correcta. No es respuesta incorrecta).

En el ítem 6, la aritmética predomina sobre las demás representaciones de los estudiantes, pero nuevamente el porcentaje de aquellos que fallaron (41%) fue mucho mayor que el de aquellos que tuvieron éxito (8%), tal como se aprecia en el Grafico 3.9 en las barras rojas. Del 15% de los estudiantes que utilizaron el diagrama de árbol el 30% respondió correctamente.

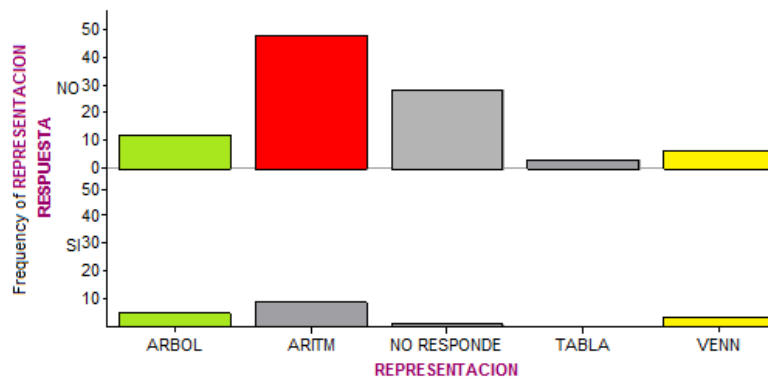


Gráfico 9. Distribución de las representaciones utilizadas en el ítem 6, prueba 2 (Si: respuesta correcta - No: respuesta incorrecta).

A Manera de conclusión se puede resaltar cómo el poco uso de las tablas o de los árboles, cuando los problemas son congruentes con ellos, es un problema o carencia que poseen los estudiantes a pesar de la enseñanza, y esto refleja que la enseñanza recibida no logró construir las estrategias válidas para resolver este tipo de problemas. Además, el uso de diagramas de Venn para problemas con probabilidades condicionales es muestra de los pobres conocimientos que los estudiantes tienen de la probabilidad condicional.

3.2.6. Otras. En este punto se presentan errores no tan generalizados, entre los que se incluyen: confusión de la condicional con su inversa, confusión de una condicional con una marginal y el valor de probabilidad mayor que 1 y se explica en qué consiste y que ítems se presentaron.

- *Confusión de la condicional con su inversa.* Otro error que también se presentó en la resolución de los problemas fue el conocido como falacia de la condicional transpuesta estudiado por varios autores como Falk, 1987; Díaz, Ortiz & Serrano, 2007; Díaz y de la Fuente, 2006. Esta falacia consiste en confundir $P(A|B)$ con $P(B|A)$. Dentro de los resultados, se encontró que en el ítem 5 de la prueba 2, en el que el 11% de los estudiantes cometió este error, porcentaje que es alto teniendo en cuenta que como ya se dijo (p. 80) fue el ítem con mayor porcentaje de no respuesta, y con otros problemas tales como: contexto, confusión de la conjunta con la condicional, propiedades de la probabilidad.

Un análisis hecho a las respuestas de los estudiantes, permitió observar, que el problema estuvo en confundir el dato que se daba como condicional (probabilidad de ojos castaños dado que tiene cabello castaño) con su inversa (probabilidad de cabello castaño dado que tiene ojos castaños). En la Evidencia 3.8, se aprecia

cómo el estudiante toma el 37,5%, que se refiere a $P(OC | CC)$ y lo asume como $P(CC | OC)$.

S. 50% → NONC
 + 5% → OC
 37,5% → OC

$$P(C/O) = \frac{P(ONC)}{P(O)}$$

$$P(O) = \frac{0,3}{37,5\%} = 8 \times 10^{-3}$$

Evidencia 3.8.

- *Confusión de una condicional con una marginal.* Otra de las dificultades que tienen los estudiantes en el momento de resolver un problema de probabilidad condicional, es confundir una probabilidad condicional con una marginal, como lo registran algunos estudios tales como los de Díaz & de la Fuente, (2007) y Díaz, Ortiz & Serrano, (2007).

En las respuestas de los estudiantes también se presentó esta dificultad y como lo muestra la Tabla 7. (p. 72) el ítem 3 fue el que reportó el mayor porcentaje (9%). En éste los estudiantes daban como respuesta la probabilidad de inteligencia superior, en lugar de dar la probabilidad de inteligencia superior dado que tiene rendimiento alto, que era lo que se preguntaba, tal como lo muestra la Evidencia 3.9.

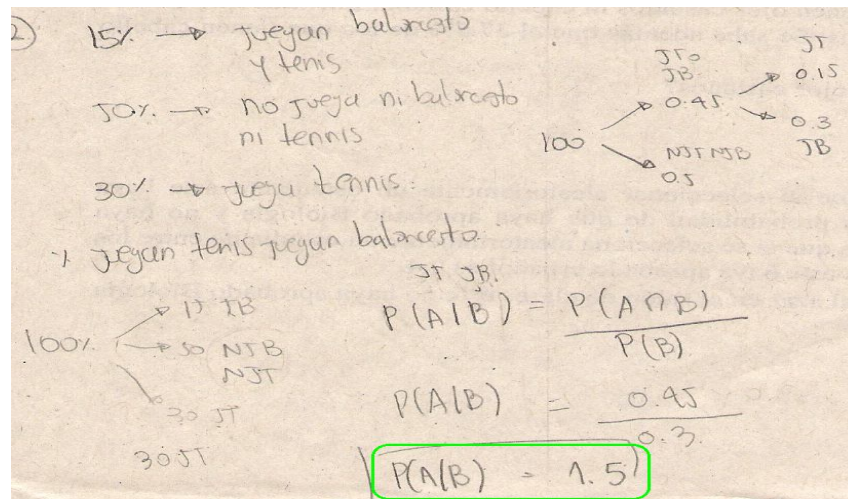
3] P. INTELIGENCIA SUPERIOR → 0,56
 P. NO TENDGA INTELIGENCIA SUPERIOR NI REND. ALTO → 0,24
 P. INTELIGENCIA SUPERIOR Y RENDIMIENTO BAJO → 0,16.

RTA = LA PROBABILIDAD ES DEL 0,56.

Evidencia 3.9.

- *Obtener un valor de probabilidad mayor que 1.* Este error se puede dar por dos razones, una es que las personas ignoran que por definición la probabilidad de un evento no puede ser mayor que 1, otra es que cuando los datos se dan como porcentajes las personas podrían pensar que el resultado que se obtiene al utilizar la definición de probabilidad condicional es otro porcentaje (que no están acotados) y no una probabilidad. La Evidencia 3.10, ejemplifica la actuación de una estudiante que no conoce la definición básica de probabilidad, pues ignora la imposibilidad de obtener probabilidades con un valor mayor a 1, en su caso 1.5. Este error a su vez está influenciado por la siguiente regla que la estudiante aplica:

$$P(A \cap B) = P(A \cap B) + P(A)$$



Evidencia 3.10.

CAPÍTULO IV

CONCLUSIONES

En esta última sección, se encuentran las conclusiones a las que se llegó después de los análisis de los resultados. En algunas se resaltan las implicaciones que tienen en la enseñanza-aprendizaje de la probabilidad condicional.

1. En el estudio del efecto del formato de presentación de un problema binario de probabilidad condicional y el tipo de información sobre las respuestas de los estudiantes, se encontró que para los problemas de caso 2 y caso 5 presentados en formatos de frecuencias naturales y probabilidad condicional, el tipo de información tiene un efecto significativo sobre las respuestas, produciendo mayor proporción de repuestas correctas aquellos en los que la información no presenta datos condicionales (nivel 0, en la clasificación de Yáñez, 2001). No obstante, estos resultados contradicen parcialmente los resultados obtenidos por Gigerenzer & Hoffrage (1995), quienes concluyeron que las frecuencias naturales mejoraban los porcentajes de respuestas correctas de manera considerable. Pero esto teniendo en cuenta que ellos trabajaron exclusivamente con problemas de tipo 7 (iii), sin embargo al tratarse en esta investigación problemas de otros tipos se concluyó que los efectos que puedan tener las frecuencias naturales (formato de presentación) no son independientes del nivel del problema que se proponga, más aún, este tipo de información es mejor cuando existe alguna información sobre probabilidades condicionales.
2. A pesar de que todos los estudiantes que participaron en esta investigación habían tomado por lo menos un curso de probabilidad, se encontraron errores que persistieron aún después de la enseñanza. Entre los más notables están la confusión de la probabilidad conjunta con la condicional y viceversa, y la falacia de la condicional transpuesta.

3. Todos los errores presentados en los ítems, llevan a plantear cambios en el proceso de enseñanza de la probabilidad condicional. Una propuesta es: en sexto y séptimo grado se enseñar al estudiante expresiones del lenguaje natural que indiquen condicionalidad, en la que ellos diferencien el dato condicionado y el condicional. En octavo y noveno grado introducirlos al lenguaje simbólico que usa la probabilidad condicional y traducir las expresiones del lenguaje natural al lenguaje simbólico. En décimo y undécimo grado presentar la fórmula de Bayes y problemas de aplicación. Finalmente en la universidad se deben formalizar todos esos conocimientos sobre probabilidad condicional, enfrentando a los estudiantes con problemas de todos los niveles y casos según la clasificación de Yáñez (2001).

4. Cuando se presenta el problema en formato de frecuencias naturales, la clasificación de Yáñez (2001), se reduce, pues en este formato existen dos tipos de datos que como número son el mismo, los que hacen referencia a un suceso conjunto y los que hacen referencia a un suceso condicional. De aquí resulta que los problemas que presentan condicionales en su información se transforman de la siguiente manera:
 - ✓ Caso 1: Tres intersecciones
 - ✓ Caso 2: Dos intersecciones y una marginal
 - ✓ Caso 3: Dos marginales y una intersección
 - ✓ Caso 4: Dos marginales y una condicional se transforma en caso 3
 - ✓ Caso 5: Dos intersecciones y una condicional se transforma en caso 1
 - ✓ Caso 6: Una marginal, una intersección y una condicional se transforma en caso 2
 - ✓ Caso 7: Una marginal y dos condicionales se transforma en caso 3
 - ✓ Caso 8: Una intersección y dos condicionales se transforma en caso 1
 - ✓ Caso 9: Tres condicionales se transforma en caso 1

Por lo tanto sólo quedarían problemas de los casos 1,2 y 3 ya que las condicionales se pueden interpretar como intersecciones.

5. La enseñanza tradicional de la probabilidad condicional, que consiste en tomar problemas que siempre requieren para su solución el teorema de Bayes o el Teorema de Probabilidad Total, han hecho que los estudiantes consideren que éstas son las únicas herramientas para resolverlos. Además, al enseñar representaciones, siempre se toman problemas de tipo 7, lo que lleva a los estudiantes a usar diagramas de árbol para resolver cualquier tipo de problema. Por todo lo anterior, muchos estudiantes construyen un algoritmo único para resolver “cualquier problema de probabilidad condicional”, sin tener en cuenta la información que presenta, lo que los lleva a fracasar en la solución de problemas cuya información contiene probabilidades conjuntas y marginales.
6. Como los problemas en formato de frecuencias permiten pasar las condicionales a intersecciones, es decir, convertir un problema de niveles 1,2, y 3 a problemas de nivel 0, y que como se vio en los resultados los problemas de este nivel producen más éxito, se debería enseñar a los estudiantes a traducir los datos de un problema en formato de probabilidades a formato de frecuencias naturales.
7. Se está de acuerdo con Lonjedo & Huerta (2007a), en la secuencia que ellos proponen para la enseñanza de la probabilidad condicional, que consiste en enseñar primero problemas en formato de frecuencias, luego en formato de porcentajes y por último en formato de probabilidades. Sin embargo, cuando el profesor hace uso del formato de frecuencias debe tener cuidado de no presentar información conjunta y condicional, porque esto podría generar en el estudiante la confusión entre éstas, por lo que él debe escoger uno de los tipos de información, o como también lo

argumentan Lonjedo & Huerta (2007a), presentar los datos condicionales en formato de porcentajes.

8. Como el lenguaje y el contexto de un problema de probabilidad condicional son factores que influyen en los procesos de solución del mismo, los profesores deben dar a sus estudiantes variedad en las expresiones para la condicionalidad, que les permitan comprender el problema. Esto exige a los maestros pensar y escoger muy bien aquellos problemas que dan a los estudiantes.
9. Las representaciones constituyen una herramienta importante a la hora de resolver un problema de probabilidad condicional y su utilidad está en saber escogerla, ya que hay representaciones apropiadas dependiendo del tipo de información que se da. El análisis de las representaciones que se hizo en el capítulo de resultados, puso en evidencia la poca capacidad de los estudiantes para discriminar entre la representación apropiada, pues con mucha facilidad utilizaron diagramas de Venn para representar información condicional. Para superar estas dificultades los profesores, deben enfatizar en los eventos que cada representación requiere en su construcción. Pero esto sólo se puede hacer cuando se enseñan diferentes tipos de problemas.
10. La discusión alrededor de las frecuencias naturales dada por Gigerenzer & Hoffrage (1995), Martignon & Wassner (2002), Lewis & Keren (1999) y, los demás, ahora debe ser ampliada considerando no sólo problemas que en su información presenten dos condicionales y una marginal, sino también problemas que den información sobre otros eventos, pues como se dijo el tipo de información tiene un efecto mayor que el que puede tener el formato de presentación para problemas que no presentan condicionales.

11. Resolver un problema binario de probabilidad condicional no es sencillo, pues en el proceso de solución intervienen factores como: la correcta identificación de los datos, la escogencia de la representación, el uso de las fórmulas correctas para responder a la pregunta y las operaciones que ésta requiere, que en el momento en que los maestros los identifican, deben buscar estrategias para evitar que éstos se conviertan en obstáculos infranqueables para los estudiantes.

REFERENCIAS

Cosmides, L. & Toby J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *COGNITION*, 58, 1-73.

Díaz, C. & de la Fuente, I. (2005). Razonamiento sobre probabilidad condicional e implicaciones para la enseñanza de la estadística. *Épsilon* 59, 245-260.

Díaz, C. & de la Fuente, I. (2006). Dificultades en la resolución de problemas que involucran el teorema de Bayes, un estudio exploratorio en estudiantes españoles de psicología. *Educación matemática: Vol. 18*, 75-94.

Díaz, C., Ortiz, J. & Serrano, L. (2007). *Un estudio experimental de las dificultades de los estudiantes en la aplicación del teorema de Bayes*. XI simposio de la Sociedad Española de Investigación en Educación Matemática. La Laguna, España, p. 199-208.

Estrada, A., Díaz, C. & de la Fuente, I. (2005). *Un estudio inicial de sesgos en el razonamiento sobre probabilidad condicional en alumnos universitarios*. X Simposio de la Sociedad Española de Investigación en Educación Matemática. En P. Bolea, M.J. González y M. Moreno (Eds.). Huesca, España, p. 277-284.

Evans, J., Handley, S., Perham, N., Over D. & Thompson, V. (2000). Frequency versus probability formats in statistical word problems. *COGNITION*, 77, 197-203.

Falk, R. (1986). *Conditional probabilities: insights and difficulties*. II International Conference on Teaching Statistics. En R. Davidson y J (Eds.). Victoria, Canadá, p. 292-297.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency format. *Psychological Review*, 102, 684-704.

Giroto, V. & González, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *COGNITION*, 78, 247-276.

Hoffrage, U., Gigerenzer, G., Krauss, S. & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *COGNITION*, 84, 343-352.

Huerta, M.P. & Lonjedo, M.A. (2003a). *La resolución de problemas de probabilidad condicional*. VII Simposio de la Sociedad Española de Investigación en Educación Matemática. En Castro, Flores et al (Eds.). Granada, España.

Huerta, M.P. & Lonjedo, M.A. (2003b). *Los problemas de probabilidad condicional en la enseñanza secundaria, en encuentro educativos*. XI Jornada para el Aprendizaje y la Enseñanza de las Matemáticas. Consejería de Educación, Cultura y Deportes del Gobierno de Canarias (Ed.). Canarias, España, p. 739-744.

Huerta, M.P. & Lonjedo, M.A. (2006). *The Nature of the quantities in a conditional probability problem. Its influence in the problem solving behavior*. IV Conference of European Research in Mathematics Education . En M. Bosch (Ed.). Sant Feliu de Guixols, España, p. 528-538.

Lewis, C. & Keren, G. (1999). On the Difficulties Underlying Bayesian Reasoning: A Comment on Gigerenzer and Hoffrage. *Psychological Review*, 106, 411-416.

Lonjedo, M.A. (2007). Análisis de los problemas ternarios de probabilidad condicional de enunciado verbal y de sus procesos de resolución. Tesis doctoral publicada en la Universidad de Valencia. España.

Lonjedo, M.A. & Huerta, M.P. (2004). *Una clasificación de los problemas escolares de probabilidad condicional. Su uso para la investigación y el análisis de*

textos. VIII Simposio de la Sociedad Española de Investigación en Educación Matemática. En Castro, de la Torre (Eds.). Coruña, España, p. 229-238.

Lonjedo, M.A. & Huerta, M.P. (2007a). *The "same" problem in three presentation formats: different percentages of success and thinking processes*. V Conference of European Research in Mathematics Education. En Demetra Pitta-Pantazi & George Philippou (Eds.). Larnaca, Cyprus, p. 732-742.

Lonjedo, M.A. & Huerta, M.P. (2007b). *Un análisis del comportamiento de los estudiantes en la resolución de problemas isomorfos de probabilidad condicional*. Investigación en educación matemática XI. En Camacho, Flores y Bolea (Eds.). La Laguna, España, p, 273-282.

Martignon, L. & Wassner, C. (2002). *Teaching Decision Making and Statistical Thinking With Natural Frequencies*. VI International Conference on Teaching Statistics. En B. Phillips (Ed.). Ciudad del Cabo, Sur África, p. 1-4.

Mellers, B. & McGraw, P. (1999). How to Improve Bayesian Reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological Review*, 106, 417-422.

Ojeda, A. M. (1995). Dificultades del alumnado respecto a la probabilidad condicional. *UNO*, 5, 37-55.

Pollatsek, A., Well, A., Konold, C. & Hardiman, P. (1987). Understanding Conditional Probabilities. *Organization Behavior and Human Decision Processes*, 40, 255-269.

Tarr, J. E. & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal*, 9, 39-59.

Tarr, J. E. & Lannin, J. K. (2005). *How can teachers build notions of conditional probability and independence?*, en G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning*. Springer, New York, 216-238.

Tversky, A & Kahneman, P. (1982). *Judgment under uncertainty: Heuristics and biases*. 84-98. New York: Cambridge University Press.

Yáñez, G. (2001). El álgebra, las Tablas y los Árboles en Problemas de Probabilidad Condicional. *Iniciación a la investigación en didáctica de la matemática. Homenaje al profesor Mauricio Castro*. En Gómez, P., y Rico, L. Granada, España, p. 355-37.

Yáñez, G. (2003). *Estudios sobre el papel de la simulación computacional en la comprensión de las secuencias aleatorias, la probabilidad y la probabilidad condicional*. Tesis doctoral no publicada. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.

Yáñez, G. (2005). *The Difficulties of the Taxi Problem: Example of Non-Congruency Between Semiotic Registers in Probability*. 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Virginia, Estados Unidos, p. 1-7.

ANEXOS

ANEXO A
PRUEBA PILOTO 1

NOMBRE: _____ **CARRERA:** _____

Resuelva los siguientes problemas mostrando el procedimiento completo que realizó para llegar a la respuesta.

- 1) La clase de matemáticas tiene 30 estudiantes entre chicos y chicas. Entre los estudiantes hay 7 chicas que usan gafas, 10 chicas que no las usan y 8 chicos que tampoco usan gafas.
 - a. Entre los chicos de la clase ¿qué porcentaje usa gafas?
 - b. Entre los que usan gafas ¿qué porcentaje son chicas?

- 2) Un centro escolar con 1000 alumnos entre niños y niñas tiene 282 estudiantes que usan gafas, 147 niñas que las usan y 368 niñas que no las usan.
 - a. Entre los niños ¿qué porcentaje usa gafas?

- 3) Una población de 30 personas con riesgo de tuberculosis se somete a un test para averiguar si padece de tuberculosis o no. El test da positivo o negativo para la enfermedad en cualquier caso. Hay 17 personas que son tuberculosas. Los resultados muestran que hay 14 personas a las que el test le ha dado positivo. Además, a 7 personas que son tuberculosas el test les ha dado negativo.
 - a. Entre las personas que no son tuberculosas ¿a qué porcentaje el test les da positivo?

- 4) Un 60% de los alumnos de un colegio aprobaron filosofía y un 70% matemáticas. Además, un 80% de los alumnos que aprobaron matemáticas, aprobaron también filosofía.

- a. ¿Qué porcentaje de alumnos aprobaron las dos asignaturas?
 - b. ¿Qué porcentaje de los que aprobaron filosofía aprobaron matemáticas?
- 5)** De todos los alumnos de un instituto, el 30% de los estudiantes practican baloncesto y fútbol, el 30% practica baloncesto y no practica fútbol, de los que no practican baloncesto el 40% practica fútbol, ¿qué porcentaje de estudiantes juega fútbol?

ANEXO B
PRUEBA PILOTO 2

NOMBRE: _____ **CARRERA:** _____

- 6)** A un grupo de 30 estudiantes de un colegio se les pregunta: ¿ha leído la novela Jurassic Park escrita por Michael Crichton?, ¿ha visto la película Jurassic Park dirigida por Steven Spielberg? De las respuestas de los estudiantes se sabe que la mitad de ellos han visto la película, pero únicamente un tercio de estos han leído la novela. Sin embargo sólo dos de los estudiantes han leído la novela pero no han visto la película.
- a. ¿Qué porcentaje de estudiantes de la clase ha leído el libro?
 - b. ¿Qué porcentaje ha leído el libro si no se ha visto la película?

- 7)** El portero titular de un equipo de fútbol para 8 de cada 10 penaltis, mientras que el suplente solo para 5. El portero suplente juega, por término medio, 15 minutos en cada partido (90 minutos).
- a. Si en un partido se lanzan tres penaltis contra este equipo, ¿cuál es la probabilidad de que se paren los tres?
 - b. Si se lanza un penalti y no se para ¿cuál es la probabilidad de que estuviera jugando el portero titular?

- 8)** En un estudio entre los chicos y chicas de bachillerato que están a favor y en contra de la prueba de acceso a la universidad se obtienen los siguientes resultados:

La probabilidad de que un chico esté en contra de las pruebas es de 0.44,

La probabilidad de que esté a favor entre los chicos es de 0.2

De las chicas la probabilidad de que esté en contra es de 0.8

Si elegido un estudiante que es chico, calcule la probabilidad de que esté en contra.

9) En un grupo de individuos se hizo un test de inteligencia y se midió su rendimiento académico. Se estudiaba la inteligencia superior y el rendimiento alto. De este estudio se sabe que:

La probabilidad de que tenga inteligencia superior sabiendo que su rendimiento es alto es $\frac{2}{3}$.

La probabilidad de que tenga rendimiento alto sabiendo que su inteligencia es superior es de $\frac{5}{7}$.

La probabilidad de que no tenga inteligencia superior sabiendo que su rendimiento no es alto es $\frac{3}{5}$.

- a. Calcule la probabilidad de tener inteligencia superior y rendimiento alto
- b. Calcule la probabilidad de tener rendimiento alto

ANEXO C

PROBLEMAS INDIVIDUALES

PROBLEMA 1: En un salón de clase con 50 estudiantes, 15 de ellos aprobaron biología y español, 18 estudiantes aprobaron biología y no aprobaron español. Se sabe además que de los que no aprobaron biología 10 no aprobaron español. Calcule la probabilidad de que elegido un estudiante al azar de los que no aprobaron biología éste haya aprobado español.

PROBLEMA 2: En un salón de clase la probabilidad de que al seleccionar aleatoriamente un estudiante éste haya aprobado biología y español es de 0.3, y la probabilidad de que haya aprobado biología y no haya aprobado español es de 0.36. Se sabe además que si se selecciona aleatoriamente un estudiante entre los que no aprobaron español la probabilidad de que éste no haya aprobado biología es 0.357. Calcule la probabilidad de que elegido un estudiante al azar de los que no aprobaron biología éste haya aprobado español.

PROBLEMA 3: En un salón de clases con 50 estudiantes, 33 aprobaron biología. 15 aprobaron biología y español. 7 aprobaron español y no aprobaron biología. Calcule la probabilidad de que elegido un estudiante al azar de los que no aprobaron biología éste haya aprobado español.

PROBLEMA 4: En un salón de clases la probabilidad de que al seleccionar un estudiante al azar éste haya aprobado biología es 0.66. La probabilidad de que haya aprobado biología y español es 0.3. Se sabe además que la probabilidad de que un estudiante haya aprobado español y no haya aprobado biología es de 0.14. Calcule la probabilidad de que elegido un estudiante al azar de los que no aprobaron biología éste haya aprobado español.

ANEXO D
PRUEBA DE ESTUDIO 1

NOMBRE: _____ **CARRERA:** _____

Resuelva los siguientes problemas mostrando el procedimiento completo que realizó para llegar a la respuesta.

- 1) Un centro escolar con 1000 alumnos entre niños y niñas tiene 282 estudiantes que usan gafas, 147 niñas que las usan y 368 niñas que no las usan.

Entre los niños ¿qué porcentaje usa gafas?

- 2) Para organizar un campeonato de verano hemos pedido a todos los inscritos sus aptitudes deportivas respecto al baloncesto y al tenis. Hemos encontrado que un 15% de los inscritos juegan baloncesto y tenis, un 50% no juega ni baloncesto ni tenis y un 30% juega tenis.

¿Qué porcentaje de los inscritos al campeonato, de los que sabemos que juegan tenis, juegan baloncesto?

- 3) En un grupo se realizó un test de inteligencia y se midió su rendimiento académico y se encontró que la probabilidad de que un individuo tenga inteligencia superior es de 0.56, además la probabilidad de que un individuo no tenga inteligencia superior ni rendimiento alto es de 0.24 y la probabilidad de que un individuo tenga inteligencia superior y rendimiento bajo es de 0.16.

Si se selecciona al azar un individuo con rendimiento alto, ¿Cuál es la probabilidad que sea superior en inteligencia?

- 4) Se sortea un viaje a Singapur entre los 120 mejores clientes de una agencia de automóviles. De ellos 45 son mujeres casadas, 35 son hombres casados y además 20 de los que no están casados son mujeres.

¿Cuál es la probabilidad de que el afortunado sea hombre?

- 5)** De todos los alumnos de un instituto, el 30% de los estudiantes practican baloncesto y fútbol, el 30% practica baloncesto y no practica fútbol, de los que no practican baloncesto el 40% practica fútbol.

¿Qué porcentaje de estudiantes juega fútbol?

- 6)** En un hotel la probabilidad de que elegido un huésped al azar éste practique tenis y golf es de 0.3 y la probabilidad de que practique tenis y no practique golf es 0.3. Si conocemos que la probabilidad de que elegido un huésped de los que no practican tenis este practique golf es 0.4.

Calcule la probabilidad de que tomado un huésped al azar no practique ni tenis ni golf.

ANEXO E
PRUEBA DE ESTUDIO 2

NOMBRE: _____ **CARRERA:** _____

Resuelva los siguientes problemas mostrando el procedimiento completo que realizó para llegar a la respuesta.

- 1) Un centro escolar con 1000 alumnos entre niños y niñas tiene 282 estudiantes que usan gafas, 147 niñas que las usan y 368 niñas que no las usan.

¿Qué porcentaje de niños usa gafas?

- 2) Para organizar un campeonato deportivo hemos preguntado a todos los inscritos si practican baloncesto y/o tenis. Hemos encontrado que un 15% de los inscritos juegan baloncesto y tenis, un 50% no juega ni baloncesto ni tenis y un 30% juega tenis.

¿Qué porcentaje de los inscritos al campeonato, **de los que** sabemos que juegan tenis, juegan baloncesto?

- 3) En un grupo de estudiantes se realizó un test de inteligencia y se midió el rendimiento académico de los estudiantes. Se encontró que la probabilidad de que un individuo tenga inteligencia superior es de 0.56, además la probabilidad de que un individuo no tenga inteligencia superior ni rendimiento alto es de 0.24 y la probabilidad de que un individuo tenga inteligencia superior y rendimiento bajo es de 0.16.

Si se selecciona al azar un individuo con rendimiento alto, ¿Cuál es la probabilidad de que tenga inteligencia superior?

- 4) Se sortea un viaje a Singapur entre los 120 mejores clientes de una agencia de automóviles. De ellos 45 son mujeres casadas, 35 son hombres casados y además 20 de los que no están casados son mujeres.

¿Cuál es la probabilidad de que el afortunado sea hombre?

- 5) En cierta ciudad el 50% de las personas no tienen ojos castaños ni cabello castaño, mientras que el 15% de ellas, tienen cabello castaño y ojos castaños. Se sabe además que el 37.5% **de los que** tienen cabellos castaño tienen ojos castaños.

¿Qué porcentaje de las personas en esta ciudad tienen ojos castaños?

- 6) En un salón de clases la probabilidad de que al seleccionar aleatoriamente un estudiante éste haya aprobado biología y español es de 0.3 y la probabilidad de que haya aprobado biología y no haya aprobado español es de 0.35. se sabe además que si se selecciona aleatoriamente un estudiante **entre los que** perdieron biología la probabilidad de que este haya aprobado español es de 0.4.

Calcule la probabilidad de que elegido un estudiante al azar en el salón de clases éste no haya aprobado biología ni español.

