

“PROBLEMÁTICAS RECONOCIDAS EN EL PROCESO DE CONSTRUCCIÓN  
DEL SIGNIFICADO DE LA ASOCIACIÓN DE VARIABLES”

SANDRA MILENA MUÑOZ AGREDO  
LUIS GUILLERMO ORTEGA PIAMBA  
ZULY MILENA VIDAL VÁSQUEZ

UNIVERSIDAD DEL CAUCA  
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN  
DEPARTAMENTO DE MATEMÁTICAS  
POPAYÁN  
2007

“PROBLEMÁTICAS RECONOCIDAS EN EL PROCESO DE CONSTRUCCIÓN  
DEL SIGNIFICADO DE LA ASOCIACIÓN DE VARIABLES”

SANDRA MILENA MUÑOZ AGREDO  
LUIS GUILLERMO ORTEGA PIAMBA  
ZULY MILENA VIDAL VÁSQUEZ

Trabajo de grado en la modalidad de seminario presentado como requisito para  
optar al título de Licenciado en Educación con Especialidad en Matemática

Director  
Magíster YILTON RIASCOS FORERO

UNIVERSIDAD DEL CAUCA  
FACULTAD DE CIENCIAS NATURALES, EXACTAS Y DE LA EDUCACIÓN  
DEPARTAMENTO DE MATEMÁTICAS  
POPAYÁN  
2007



**NOTA DE ACEPTACIÓN**

---

---

---

---

---

Magister. Yilton Riascos Forero.  
**Director**

---

Especialista Maria del Pilar Astudillo  
**Evaluador**

Magíster Edwin Rengifo Canizales  
**Evaluador**

Popayán, 12 de junio de 2007

## DEDICATORIA

*A Dios, por llamarnos a su misma misión.*

*“...Por eso suplique y se me concedió la prudencia; invoqué y vino a mi el espíritu de sabiduría.*

*La preferí a cetros y tronos y en su comparación tuve en nada la riqueza.*

*No la equiparé a la piedra mas preciosa, por que todo el oro a su lado es un puñado de arena y ante ella la plata es como el barro...”*  
*Sabiduría 7: 7-9*

*A quienes debemos la respuesta ¿Cuándo se van a graduar?*

*A las bendiciones hechas realidad aquí en la tierra, nuestros padres*

*Edgar Muñoz y Maria Agredo; Efraín Vidal y Maria Rita Vasquez;  
Luis Guillermo Ortega y Amparo Piamba.*

*A los cultivados en la misma tierra: Juan Pablo, Andrea, Martha, Alejandra,  
y Eduardo Muñoz Agredo; Gueldy Vidal Vasquez; Lisbeth Amparo Ortega  
Piamba,*

## **AGRADECIMIENTOS**

A nuestro dilecto director YILTON RIASCOS.

Por su entrega, sacrificio, paciencia, amistad sincera y visión para dirigir este barco, de inexpertos marineros

Por educarnos con el ejemplo, en la solidaridad, tolerancia y justicia.

A la Especialista Maria del Pilar Astudillo, y al Magíster Edwin Rengifo Canizales, miembros del Comité de Evaluación y Seguimiento.

Al gran taller llamado Universidad del Cauca, quien nos recibió como arcilla, nos moldeó y ahora nos devuelve a la sociedad como piezas templadas para ser pulidas de poco en poco por nuestro quehacer.

## CONTENIDO

	Pág
1. PRESENTACIÓN	1
2. INTRODUCCIÓN	4
3. JUSTIFICACIÓN	6
4. OBJETIVOS	9
4.1. OBJETIVO GENERAL	9
4.2. OBJETIVOS ESPECÍFICOS	9
5. METODOLOGÍA	10
6. GENERALIDADES DEL TERMINO ASOCIACIÓN	12
6.1. ALGUNAS CONCEPCIONES.	12
6.2. ASOCIACIÓN ESTADÍSTICA	12
7. DISTRIBUCIONES BIDIMENSIONALES DE FRECUENCIA	16
8. ALGUNAS MEDIDAS DE ASOCIACIÓN	22
8.1. COEFICIENTES DE CORRELACIÓN.	22
8.1.1. Coeficiente de Correlación Lineal o r de Pearson	22
8.1.2. Cuadrado de contingencia o Coeficiente chi cuadrado de Pearson ( $\chi^2$ ).	24
8.1.3. Coeficiente Phi ( $\Phi$ )	26
8.1.4. Coeficiente de Correlación de Cramer (V)	26
8.1.5. Coeficiente de Contingencia (C)	27
8.1.6. Coeficiente de Correlación de Spearman ( $\rho$ )	27
8.2. TABLAS DE CONTINGENCIA	28
8.3. INDEPENDENCIA ESTADÍSTICA	30
8.4. CORRELACIÓN Y CAUSALIDAD	31
9. INCORPORACIÓN DE LA ESTADÍSTICA EN LA ESCUELA.	33
10. UNA PROPUESTA TEÓRICA PARA EL ANÁLISIS DIDÁCTICO	42
11. LA ASOCIACIÓN ESTADÍSTICA DESDE LA PERSPECTIVA DIDÁCTICA	45
11.1. INVESTIGACIONES DIDÁCTICAS DESDE LA PERSPECTIVA PSICOLÓGICA	47
11.2. INVESTIGACIONES DIDÁCTICAS A PARTIR DE LA DIDÁCTICA DE LA MATEMÁTICA	49
12. RESULTADOS DE LAS PROBLEMÁTICAS	53
12.1. CONSTRUCCIÓN DEL SIGNIFICADO DE LA ASOCIACIÓN ESTADÍSTICA MEDIANTE EL ANÁLISIS DE DATOS	54
12.2. CORRELACIÓN Y REGRESIÓN EN LSO PRIMEROS CURSOS UNIVERSITARIOS	59

13. CONCLUSIONES	65
14. BIBLIOGRAFÍA	67
15. ANEXOS	69

## 1. PRESENTACIÓN

Habiendo culminado los cursos regulares del currículo del programa de Licenciatura en Educación con Especialidad en Matemáticas y preocupados por adelantar nuestro trabajo de grado, encontramos que la Resolución 197 de 2002 emanada del Consejo de la Facultad de Ciencias Naturales, Exactas y de la Educación (FACNED), plantea las alternativas posibles para cumplir el requisito de trabajo de grado para optar al título; una de las cuales consiste en realizar un seminario de grado en la modalidad de seminario especial, que es por el que hemos optado.

Aprovechando que durante nuestra formación académica asistimos al curso de Didáctica de la Matemática y además, que las asignaturas del área electiva fueron desarrolladas en Estadística, surgió el interés de realizar un trabajo que permitiera abordar una problemática que asociara el trabajo con objetos de la estadística en el ámbito escolar, llegando a pensar en el tema de la enseñanza de la asociación de variables, que es interés de la Didáctica de la Estadística, y que en la estructura académico administrativo del Departamento de Matemáticas se enmarca en el trabajo que, sobre Pedagogía de las Matemáticas, adelanta el Grupo de Estudio y Desarrollo Investigativo en Filosofía e Historia de la Matemática y Educación Matemática.

La Estadística es una disciplina relativamente nueva en comparación con la matemática, y si hablamos de la Didáctica de la Estadística, estamos incursionando en un área que prácticamente es desconocida para la gran mayoría de los profesionales, como lo hace notar Batanero (2001), al exponer los siguientes factores:

- Existe un grupo muy reducido de personas que conocen acerca del tema.
- Es poca la investigación que en esta área se ha desarrollado.
- Es menor aún la publicación y por ende difusión de los contenidos acerca de la Didáctica de la Estadística.

Consideramos entonces necesario, ahondar en proyectos que vayan encaminados en el estudio e investigación dentro del contexto de la Didáctica de la Estadística, específicamente en las problemáticas abordadas en la enseñanza de las diferentes técnicas estadísticas, y particularmente en el tema de la asociación de variables que se enseña en las ciencias que requieren de la colaboración de los métodos de la disciplina estadística.

La medida de la asociación de variables es un tema de gran importancia en la Estadística, ya que trata el problema de cómo explicar el comportamiento de una variable dependiente, a partir del conocimiento de los valores de una o más variables independientes, es decir permite determinar la magnitud de la relación entre las variables. Este tema resulta útil en muchos campos de aplicación como: Economía, Administración, Ingenierías, Ciencias de la Salud, entre otros, para

ayudar a la interpretación de los fenómenos que ocurren de manera tal que puede esperarse algún tipo de relación entre ellos.

En este seminario se pretende identificar problemáticas que desde las investigaciones en Didáctica de la Estadística se reconocen en el proceso de construcción del significado de la asociación de variables, para ser difundidas posteriormente a estudiantes universitarios, profesionales y organizaciones interesadas en el tema, dado que estos son entes receptores y potenciales críticos de este trabajo.

## 2. INTRODUCCIÓN

El hombre como ser social, permanentemente se encuentra expuesto a toda clase de información del mundo y en particular a la de las diversas ciencias. De la lectura e interpretación que hace de ésta, depende en gran medida la toma de decisiones en forma acertada, hecho para el cual no todos estamos capacitados, dada la carencia de una cultura estadística que permita adquirir y desarrollar la capacidad para evaluar e interpretar críticamente la información que en términos estadísticos se está produciendo, para superar esta dificultad los investigadores en educación estadística revelan la necesidad de incorporar la estadística al currículo desde la escuela, ya que consideran que ésta es una de las forma mas seguras de llegar a gran parte de los ciudadanos.

Un componente de esta cultura estadística radica en educar en los conceptos básicos entre los cuales se encuentra el objeto asociación estadística, que resulta ser de gran importancia. Este objeto emerge de diferentes campos de problemas que se presentan en diversas ciencias como salud, ingeniería, economía, administración, finanzas, sociología, estadística, psicología, y educación matemática. Es en estas dos últimas, donde se registran trabajos alrededor de la comprensión de la asociación y del significado de los objetos matemáticos, los cuales sirven como apoyo a las investigaciones concernientes, interesadas en

identificar problemáticas que se generan en la construcción del significado del objeto asociación estadística.

Las problemáticas que se reconocen en este trabajo de grado, se relacionan en experimentos que se basan en la teoría construida por el equipo de investigación en educación matemática de la Universidad de Granada en cabeza del profesor Juan Díaz Godino. Tal teoría, aporta elementos precisos del significado de los objetos matemáticos, con los cuales se hace posible detectar el estado de apropiación del concepto por parte de un individuo, respecto del significado aceptado por la comunidad, en este caso, la estadística. No obstante se hace necesario abordar desde lo disciplinar los conceptos, asociación estadística, medidas de asociación, distribución de variables bidimensionales, independencia estadística, tablas de contingencia y causalidad y correlación, que se consideraron relevantes para el desarrollo de este seminario.

### 3. JUSTIFICACIÓN

Todos los espacios de formación curricular son un ambiente propicio para generar conocimiento, ya sea de tipo empírico, práctico o teórico; y es en la universidad donde a través de las distintas prácticas y modalidades de trabajos de grado, donde los estudiantes incursionan en la investigación. Aquí el estudiante se encuentra frente a problemas que necesitan ser estudiados, modelados, resueltos e interpretados, es decir problemas abordados metodológicamente.

Como lo expresan Rodríguez de Villaquirán y otros (1997) “La estadística como disciplina tiene como propósito suministrar información con base en la obtención de datos a fin de diagnosticar y tomar decisiones para la solución de problemas” los cuales surgen al querer comprender hechos que suceden a nuestro alrededor. La matemática junto con otras ciencias han permitido en gran parte reconocer el mundo que nos rodea, pero existen, fenómenos que no son fácilmente modelables o comprensibles ya que no responden a patrones conocidos y no es fácil encontrar o diseñar uno que los represente, como es el caso de los fenómenos aleatorios. Para resolver este problema surge dentro de la estadística, la inferencia estadística, rama que con un conjunto de métodos y utilizando el mayor grado de verosimilitud, pretende dar respuestas válidas y concretas.

Cuando nos vemos abocados a solucionar problemáticas relacionadas con la educación y en particular con la enseñanza de la estadística, es preciso recurrir a la Didáctica de las Ciencias, en este caso la Didáctica de la Estadística, que en analogía con la Didáctica de las Matemática, como lo expresa Godino y otros (1994) es una disciplina en formación que estudia los procesos de enseñanza/aprendizaje de los saberes estadísticos en los procesos teóricos conceptuales y de resolución de problemas tratando de caracterizar los factores que condicionan dichos procesos.

La medida de la asociación de variables es una herramienta muy útil en todo cuanto tiene que ver con encontrar relaciones entre las variables que afectan un fenómeno; y saber en que grado estas variables se relacionan resulta de vital importancia para la interpretación del mismo. Encontrar estos patrones de comportamiento y relaciones entre variables es indispensable para poder predecir hechos que sucederán a futuro (corto o mediano plazo) ya sea en el contexto social, económico, político o científico.

Dada la finalidad de este seminario, es importante que como futuros profesionales de la educación se busque estar al tanto de estos resultados y específicamente en los que tienen que ver con la didáctica ya que a través de los años se han implementado modelos psicopedagógicos buscando mejores resultados en la apropiación de los conceptos y significados, y es una ventaja conocerlos para procurar mejorar la forma de enseñarlos.

Al ser una comunidad minoritaria la que esta trabajando en el campo de la Didáctica de la Estadística, este trabajo adquiere relevancia al permitir que se genere difusión y que se procure abrir espacios de discusión sobre estas temáticas.

Este trabajo busca como logros, además de ayudarnos a profundizar en los temas de Asociación de variables, Correlación y Regresión lineal, tanto desde su concepción epistemológica como de sus aplicaciones, generar conocimiento acerca de la investigación en Didáctica de la Estadística que es un área que hasta el momento nos atrevemos abordar, además esperamos nos permita ser facilitadores en el proceso de consulta y referencia para otras persona que estén interesadas en temas afines.

## **4. OBJETIVOS**

### **4.1. OBJETIVO GENERAL**

- Identificar problemáticas que se reconocen en el proceso de construcción del significado de la asociación de variables, a partir de las investigaciones desarrolladas en Didáctica de la Estadística.

### **4.2. OBJETIVOS ESPECÍFICOS**

- Estudiar los conceptos fundamentales de la asociación de variables y de la Didáctica de la Estadística.
- Realizar una revisión bibliográfica de artículos, documentos y textos en los cuales se aborden problemáticas acerca de los procesos de enseñanza y aprendizaje de la asociación de variables.
- Caracterizar problemáticas que se reconocen en el proceso de construcción del significado de la asociación de variables y que son presentadas en los resultados de las investigaciones en Didáctica de la Estadística.
- Divulgar los resultados obtenidos de este seminario a través de charlas, presentaciones y asesorías a quienes estén interesados.

## 5. METODOLOGÍA

Acceder al conocimiento es un proceso teórico–práctico, donde las ideas e hipótesis deben ser confrontadas permanentemente con los sucesos (o hechos) para poder afirmarlas o negarlas. Es por ello que se considera importante trazar un modelo conceptual y operativo que permita cumplir el propósito del trabajo.

Ya que no existe un camino preestablecido para el manejo de la información bibliográfica, se señalarán a continuación las tareas básicas que se cumplirán:

La etapa inicial consistirá en capacitarnos y apropiarnos los conceptos fundamentales de la estadística y la didáctica de la Estadística pertinentes para el desarrollo del seminario, y posteriormente en realizar una “Revisión Bibliográfica” que consistirá en conocer, explorar y estudiar todo conjunto de fuentes que ofrezca utilidad tales como: libros, revistas, artículos, tesis de maestría y doctorado, y en general documentación que nos ayudará a desarrollar cada uno de los temas acerca de las problemáticas abordadas en la enseñanza y el aprendizaje de la asociación de variables.

Se hará una lectura discriminatoria de los documentos, teniendo en cuenta la pertinencia del contenido y el autor que elabora el documento.

Se realizarán fichas y se ordenarán según los temas tratados a medida que se realicé el estudio detallado de estos temas, esto permitirá ir desarrollando el cuerpo del documento final.

Para la elaboración del documento final se compararán las fichas, se establecerán características entre los resultados de las investigaciones en Didáctica de la Estadística que abordan problemáticas en la construcción del significado de la asociación de variables, y se elaborarán las conclusiones de todo el proceso realizado.

Por último se divulgarán los resultados obtenidos en este seminario a través de charlas, presentaciones y conversatorios con quienes estén interesados.

## 6. GENERALIDADES DEL TERMINO ASOCIACIÓN

### 6.1. ALGUNAS CONCEPCIONES.

Es frecuente escuchar hablar de asociación, este término posee diversos significados; por ejemplo, cuando se habla de asociarse para formar una empresa, se refiere a la acepción<sup>1</sup> “Acción y efecto de asociar o asociarse” o “Conjunto de los asociados para el mismo fin y, en su caso, persona jurídica por ellos formada”; cuando se trata de ligar una característica a un grupo, se entiende como<sup>2</sup> “Figura que consiste en decir de muchos lo que solo es aplicable a varios o a uno solo, ordinariamente con el fin de atenuar el propio elogio o la censura de los demás”, sin embargo, para los propósitos de este seminario es de interés el significado desde el punto de vista de la asociación de ideas que se concibe como<sup>3</sup> “*conexión mental entre ideas, imágenes o representaciones por su semejanza contigüidad o contraste*”, es en esta última definición donde se enmarca el concepto de asociación estadística que se tratará en adelante.

---

<sup>1</sup> REAL ACADEMIA ESPAÑOLA, 1992.

<sup>2</sup> Ídem

<sup>3</sup> Ídem

## 6.2. ASOCIACIÓN ESTADÍSTICA

El mundo que nos rodea presenta gran diversidad de fenómenos, de los cuales algunos podemos entender, y existen aquellos que no entendemos ni les encontramos una explicación, esto se debe a que la ocurrencia de un fenómeno depende, posiblemente, de varios factores y/o variables, y el saber de que modo o en que magnitud están relacionados es determinante para la interpretación o pronóstico del mismo.

La estadística, en el nivel de análisis de datos, se encarga de recolectar, integrar, procesar e interpretar la información, y es en este proceso en donde se detecta la existencia o ausencia de una o mas relaciones entre las variables, (e interesa medir su intensidad). A esta idea intuitiva le corresponde el concepto de *asociación estadística*.

En la literatura se encuentran diversas definiciones del término asociación, en este seminario se acoge la definición propuesta por Batanero (2001), quien define la asociación estadística como “*la medida de la intensidad de la relación entre dos variables*”, para el caso general, y hace énfasis en hablar de correlación cuando se trata con variables de tipo numérico.

Como lo expresan Runyon y Haber (1992), para abordar el estudio de la relación entre las variables que afectan un fenómeno, es preciso conocer la naturaleza de dichas variables, así para el caso en que las variables son de tipo cuantitativo se

estudiará la correlación, si son cualitativas se estudiará la asociación mediante el análisis de tablas de contingencia, y en el caso que se de una combinación de los tipos de variables anteriormente mencionadas, es decir mixtas, existen varios procedimientos, sin embargo no existe uno predefinido exactamente, podemos mencionar algunos como: comparación de la diferencia entre las medias, o el test de diferencia de medias.

El tipo de escala en el cual están medidas las variables es un factor importante para la determinación del grado de asociación entre ellas, como lo expresa Runyon y Haber (1992), ya que dependiendo de ésta escala se mide la relación de las variables utilizando diferentes coeficientes de correlación como se observa mas adelante en la tabla No. 1.

Una vez se ha percibido la existencia de una relación entre las variables, puede ser de interés encontrar un modelo matemático que permita predecir una variable en función de la otra. Cuando se van a modelar fenómenos en los que intervienen variables de tipo cuantitativo, lo que se hace es estudiar la asociación cuantitativa (correlación), con el objetivo de determinar la intensidad en que están relacionadas o la forma en que tienden a variar conjuntamente dichas variables. La expresión cuantitativa del grado de relación viene dada en términos de la magnitud del coeficiente de correlación apropiado, más adelante se describen algunos coeficientes.

En la siguiente tabla se muestran los diferentes tipos de escala en los cuales se miden las variables, el tipo de coeficiente que se debe utilizar para determinar una posible relación dependiendo de la naturaleza de las variables.

**Tabla 1. CLASIFICACIÓN DE LOS COEFICIENTES DE CORRELACIÓN SEGÚN LA ESCALA DE MEDIDA DE LAS VARIABLES.**

ESCALA	COEFICIENTE	SE USA CON
Nominal	$r_{phi}$ (coeficiente phi)  $r_b$ (r biseral)  $r_t$ (r tetracórica)	Dos variables dicotómicas.  Una variable dicotómica cuya continuidad se admite y una variable que puede tomar más de dos valores  Dos variables dicotómicas cuya continuidad se puede admitir.
Ordinal	$r_s$ (r de Spearman)  $\tau$ (tau de Kendall, o coeficiente de correlación por rangos)	Datos ordenados según su rango. Si una variable es propiamente ordinal y la otra es de intervalo/razón, se las debe expresar a las dos según su rango antes de calcular el coeficiente.  Datos ordenados según su rango.
Intervalo/razón	$r$ (de Pearson)	Dos variables de valores continuos.

Fuente: Tomado de Runyon y Haber (1992)

## 7. DISTRIBUCIONES BIDIMENSIONALES DE FRECUENCIA

En la vida cotidiana se presentan numerosos ejemplos en los cuales es de interés realizar el tratamiento conjunto de dos características o variables observadas en los elementos de una muestra o de una población, en estas situaciones puede interesar al investigador, conocer la distribución conjunta de frecuencias de los diferentes valores (o categorías) de las variables, considerar el comportamiento estadístico de una variable para los elementos que tienen un valor determinado en la otra variable, en otros casos puede interesar explorar el grado de asociación de las dos características en la población, o para predecir el valor de una de las características aprovechando el conocimiento del valor de la otra.

Al realizar un estudio es preciso determinar la naturaleza de las variables debido a que pueden ser atributos (variables cualitativas), en otros una de ellas es cuantitativa y la otra cualitativa o ambas son cuantitativas.

En general se usará la notación  $x_1, x_2, \dots, x_i, \dots, x_m$  para representar las  $m$  categorías a considerar para clasificar los elementos de la muestra (o de la población) en lo que respecta a la variable  $X$ . Estas categorías o modalidades pueden corresponder a nombres si se trata de variables medidas en escala nominal

(variables cualitativas); pueden coincidir con los valores que toma la variable  $X$ , si es cuantitativa, o pueden representar los límites de los intervalos de clase (o la marca de clase del intervalo), si  $X$  es una variable numérica continua.

Análogamente  $y_1, y_2, \dots, y_i, \dots, y_s$  representan las  $s$  categorías a considerar para clasificar los elementos de la muestra (o de la población) en lo que respecta a la variable  $Y$ . Cuando los elementos de una muestra se clasifican simultáneamente por dos características  $X$  e  $Y$ , surge, para su representación, la “tabla de doble entrada”.

En general, se extrae de la población en estudio una muestra con el propósito de observar las variables  $X$  e  $Y$ ; los datos obtenidos se presentan en parejas de la forma  $(x_i, y_j)$ , con  $1 \leq i \leq m$  y  $1 \leq j \leq s$ . Donde  $i, j, m, s \in \mathbb{N}$ .

Se entiende que un elemento de la muestra (población) se clasifica en sólo una categoría de  $X$  y en sólo una categoría de  $Y$ .

Si se llama  $\Omega$  al conjunto de todos los elementos de la muestra (población) y se llama  $X_i$  al conjunto de elementos de la muestra que posee la característica  $x_i$ , y  $Y_j$  al conjunto de elementos de la muestra que posee la característica  $y_j$ .

Para variables discretas, estas parejas se clasifican y se cuentan en una tabla de distribución conjunta de frecuencias, donde las frecuencias que aparecen en los

márgenes de la tabla son llamadas frecuencias marginales y dan cuenta individualmente, del número de elementos de la población que posee una de las características; y en forma completa, de la distribución de frecuencias de la variable.

**Tabla 2 DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS DE DOS VARIABLES.**

DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS ABSOLUTAS DE LAS VARIABLES

$X \backslash Y$	$y_1$	...	$y_j$	...	$y_s$	
$x_1$	$n_{11}$	...	$n_{1j}$	...	$n_{1s}$	$n_{1*}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$n_{i1}$	...	$n_{ij}$	...	$n_{is}$	$n_{i*}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_m$	$n_{m1}$	...	$n_{mj}$	...	$n_{ms}$	$n_{m*}$
	$n_{*1}$	...	$n_{*j}$	...	$n_{*s}$	$n_{**}$

DISTRIBUCIÓN CONJUNTA DE FRECUENCIAS RELATIVAS DE LAS VARIABLES

$X \backslash Y$	$y_1$	...	$y_j$	...	$y_s$	
$x_1$	$h_{11}$	...	$h_{1j}$	...	$h_{1s}$	$h_{1*}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$h_{i1}$	...	$h_{ij}$	...	$h_{is}$	$h_{i*}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_m$	$h_{m1}$	...	$h_{mj}$	...	$h_{ms}$	$h_{m*}$
	$h_{*1}$	...	$h_{*j}$	...	$h_{*s}$	1.0

El dato  $n_{ij}$  da cuenta del número de elementos de la muestra (población) que posee la característica  $x_i$  de la variable  $X$  y la característica  $y_j$  de la variable  $Y$  en forma simultánea, y  $h_{ij} = \frac{n_{ij}}{n}$  representa la frecuencia relativa equivalente; es decir el porcentaje de la muestra (población) que posee las dos características  $x_i$  y  $y_j$ , y los  $n_{i*}$  dan cuenta del número de elementos de la muestra que pertenece a la categoría  $X_i$ , o poseen la característica  $x_i$ , es decir para las cuales  $X = x_i$ .

Los  $h_{i*}$  indican la fracción o porcentaje del total de elementos de la muestra, que pertenecen a la categoría  $X_i$ , o que posee la característica  $x_i$ , es decir, que para el  $h_{i*}$  por ciento de la muestra se cumple  $X = x_i$ .

Los  $n_{*j}$  dan cuenta del número de elementos de la muestra que pertenece a la categoría  $Y_j$ , o posee la característica  $y_j$ , es decir, para las cuales  $Y = y_j$

Los  $h_{*j}$  indican la fracción o porcentaje del total de elementos de la muestra, que pertenecen a la categoría  $X_i$ , o posee la característica  $y_j$ , es decir, que para el  $h_{*j}$  por ciento de la muestra se cumple  $Y = y_j$ , que representan las frecuencias marginales.

Para el caso de variables continuas, se establecen categorías para cada variable y se representan por los límites de los intervalos de clase. La manera de proceder, regularmente, es la siguiente: después de tomar los datos de la muestra (población), se construyen los intervalos y la tabla de distribución conjunta de frecuencias absolutas  $n_{ij}$  y relativas  $h_{ij}$ , aquí también se pueden construir las distribuciones marginales de frecuencias para cada una de las variables.

En cuanto a la interpretación se debe tener en cuenta que estos intervalos no son directamente comparables ya que la amplitud de estos no necesariamente es igual; para solucionar este inconveniente, se plantea la estandarización de las frecuencias relativas, definiendo el concepto de densidad por unidad de área

$h^*_{ij} = \frac{n_{ij}}{A_{ij}}$  , donde  $A_{ij}$  = área de la región determinada por  $x_i \cap y_j$  , con el supuesto

de que los datos están uniformemente distribuidos.

En el caso de una variable continua y otra discreta, esta última se clasifica por categorías y para la variable continua se construyen los intervalos, los datos de la muestra (población) se clasifican y cuentan, dando origen a las tablas de distribución conjunta absoluta y relativa.

En ocasiones es de interés la distribución de frecuencias relativas de una característica, pero sólo para los elementos que satisfacen cierta condición, lo que

se conoce como frecuencias condicionales y se notan  $h(x_i / y_j) = \frac{h(x_i \cap y_j)}{h(y_j)}$  , de

esta definición puede deducirse de forma intuitiva, lo correspondiente a la regla de la multiplicación, como

$$h(x_i / y_j) = \frac{h(x_i \cap y_j)}{h(y_j)} \text{ y } h(y_j / x_i) = \frac{h(x_i \cap y_j)}{h(x_i)} \text{ se tiene que}$$

$h(y_j)h(x_i / y_j) = h(x_i \cap y_j) = h(x_i)h(y_j / x_i)$  esto es, la distribución conjunta de frecuencias relativas puede escribirse como el producto de la distribución marginal de una de las variables multiplicada por la distribución condicional de frecuencias de la otra.

Al realizar interpretaciones de frecuencias en estas tablas, se debe tener en cuenta:

Precisar el conjunto sobre el cual se esta haciendo la afirmación, ya que puede ser respecto a la población total  $n = n_{**}$  o sobre una de las categorías  $X_i, Y_j$  en que se fracciona la población, es decir se puede interpretar una frecuencia condicional como si fuera la frecuencia relativa.

La asociación estadística no guarda una relación causa-efecto, por esta razón, no debe atribuirse la existencia de una relación proporcionado por el alto porcentaje de elementos sobre una de las categorías.

Cuando se va a generar una conclusión que involucre las dos características se debe hacer sobre el total de la población y no comparando las distribuciones condicionales respectivas. Como lo expresan Behar y Yepes (1992).

## 8. ALGUNAS MEDIDAS DE ASOCIACIÓN

### 8.1. COEFICIENTES DE CORRELACIÓN.

#### 8.1.1. *Coefficiente de Correlación Lineal o r de Pearson*

Este coeficiente fue construido bajo el supuesto de que los datos tienen una distribución normal bivariada, mide el grado asociación lineal entre las dos características y es de utilidad cuando la escala de medición es de intervalo/razón, aunque tiene como ventaja que es invariante ante cambios de escala en la medición de las variables.

En Canavos (1988) y Runyon y Haber (1992) se presentan las propiedades del coeficiente de correlación, se demuestra que su valor se encuentra entre  $-1$  y  $+1$ , ambos representan relaciones perfectas entre las variables (el primero inversa y el segundo directa), y el  $0$  representa ausencia de relación lineal, lo que mas adelante denotaremos como independencia estadística de dos variables.

Es de saber que dos variables pueden estar muy relacionadas en forma no lineal (curvilínea, exponencial, potencial, logarítmica, polinómica), y tener un coeficiente

de correlación muy alto, pero al medir el coeficiente de correlación lineal este será muy bajo o tendiente a cero, dado que la relación no es lineal.

En Castillo y otros (1994) se presentan otras características de este coeficiente, las cuales hacen referencia a los que se entiende por relación lineal positiva, o directa: cuando un incremento o decremento en una de las variables genera un incremento o decremento respectivamente en la otra variable, si la correlación es cercana a uno se dice que las variables son concordantes.

Se habla de una relación negativa o inversa cuando un incremento en una de las variables tiende a generar un decremento en la otra variable o viceversa, si la correlación es cercana a menos uno se dice que existirá discordancia.

Los valores del coeficiente cercanos a cero indican que a algunos valores grandes o pequeños de  $X$  se corresponden a con valores pequeños y grandes de  $Y$ . Es de notar también que un coeficiente diferente de cero no asegura correlación entre las variables, esto último solo se probará efectuando una prueba de hipótesis, también es de resaltar que el coeficiente de correlación  $r$  por si mismo no puede probar o refutar la causalidad de una relación entre  $X$  y  $Y$ , aún si el coeficiente indica una relación perfecta ( $r=1$  ó  $r=-1$ ) Canavos (1988) propone que para calcular el coeficiente, se proceda de la siguiente forma:

Se supone la existencia de un conjunto de  $n$  mediciones  $y_1, y_2, \dots, y_s$  de una variable respuesta  $Y$ , las cuales se han observado bajo un conjunto de condiciones experimentales  $x_1, x_2, \dots, x_m$  que representan los  $m$  valores de predicción.

Se extraen de los datos, las características principales de una relación que no es evidente, se calcula el coeficiente de correlación, para este caso el  $r$  de Pearson,

mediante la fórmula  $r = \frac{\sum_{k=1}^n (X_k - \bar{X})(Y_k - \bar{Y})}{S_x S_y}$ , donde  $X$  e  $Y$  representan las variables

medidas  $\bar{X}$ ,  $\bar{Y}$  las medias y  $S_x$ ,  $S_y$  las desviaciones.

Cuando una o ambas variables  $X$  y  $Y$  sólo se miden en escala nominal, se puede obtener alguna de las medidas de asociación que se mencionan a continuación, ya que se pueden obtener a partir de una tabla de doble entrada.

### 8.1.2. Cuadrado de contingencia o Coeficiente chi cuadrado de Pearson ( $\chi^2$ ).

Behar y Yepes (1992), presentan el coeficiente Chi cuadrado, cuadrado de contingencia o Chi cuadrado de Pearson como una medida de la asociación que se denota por:

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^s \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*}$$

donde  $n_{ij}$  es la frecuencia observada y  $n_{ij}^*$  es la frecuencia esperada bajo la hipótesis de independencia  $\left( n_{ij}^* = \frac{n_{i*} \times n_{*j}}{n_{**}} = n_{ij} \right)$ ; cuando los valores de las

frecuencias observadas y esperadas coinciden se tiene  $\chi^2 = 0$ , lo cual indica que las variables son independientes, más aún,  $\chi^2$  crece cuando crecen las diferencias entre las frecuencias observadas y esperadas, un inconveniente del coeficiente, es que está afectado por el número de observaciones  $n$ , lo cual no es conveniente puesto que el grado de dependencia debe medir la diferencia entre  $\frac{n_{ij}}{n}$  y  $\frac{n_{i*}}{n} \times \frac{n_{*j}}{n}$ , que como puede apreciarse no varia si se multiplican todas las

casillas  $n_{ij}$  por una constante  $c$ , lo cual es equivalente a multiplicar el número de observaciones por ese mismo factor; este aspecto puede corregirse definiendo el “cuadrado medio de contingencia  $f^2$ ”, que se calcula por

$$f^2 = \frac{\chi^2}{n} = \left[ \sum_{i=1}^m \sum_{j=1}^s \frac{n_{ij}^2}{n_{i*} n_{*j}} - 1 \right],$$

es siempre mayor que cero y no esta acotado en forma general, pero tiene una cota superior para cada problema en función del número de categorías de  $X$  y  $Y$ .

### 8.1.3. Coeficiente Phi ( $\Phi$ )

Según Castillo (1994), el coeficiente *Phi* es una estandarización de  $\chi^2$  y está dado por:

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

Este estadístico toma el valor 0 si no hay asociación y 1 si todas las frecuencias ocurren con la diagonal principal de la tabla. Es una medida apropiada para las tablas de dimensión 2 x 2 que se forman a partir de variables medidas en escala nominal.

### 8.1.4. Coeficiente de Correlación de Cramer (V)

En Behar y Yepes (1996) y Castillo y otros (1994) se muestra éste coeficiente como una estandarización de  $\chi^2$ , cuyo rango no está afectado por el número de categorías en  $X$  y  $Y$ ,  $0 \leq V \leq 1$ , es apropiado en escala nominal y se define como :

$$V = \sqrt{\frac{\chi^2}{n \min\{(s-1), (m-1)\}}}$$

$V^2 = 0$  Indica independencia de las variables y  $V^2 = 1$  indica el máximo grado de dependencia es decir el comportamiento de una variable explica exactamente el comportamiento de la otra, lo cual permite medir la intensidad de la dependencia.

### 8.1.5. Coeficiente de Contingencia ( $C$ )

En Castillo y otros (1994) y Siegel (1991) se encuentra que este coeficiente se puede ver como otra estandarización de  $\chi^2$ , y mide el grado de asociación entre dos conjuntos de atributos. Es útil cuando solo se tiene información clasificatoria (escala nominal) acerca de uno o ambos conjuntos de atributos.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n_{**}}}$$

Esto es, puede usarse cuando la información acerca de los atributos consiste en una serie no ordenada de frecuencias. El coeficiente de contingencia varía entre  $0 \leq C \leq K$ , donde  $C = 0$ , indica una completa carencia de asociación, y nunca alcanza la unidad (1), el límite superior  $K$  para este coeficiente es una función del número de filas y columnas de la tabla  $n_{XS}$ , esto hace que sea más difícil su interpretación.

### 8.1.6. Coeficiente de Correlación de Spearman ( $\rho$ )

Siegel (1991) presenta este coeficiente como una prueba que mide la asociación entre dos variables basándose en rangos, además permite desarrollar pruebas de hipótesis, es una medida que requiere que ambas variables sean medidas por lo menos en una escala ordinal. Para calcular  $\rho$ , se hace una lista de las  $n$  observaciones, después de registradas, se anotan los rangos en la variable  $X$  y

$Y$ , se determinan los distintos  $d_k$  que son los valores de la diferencia entre los correspondientes valores  $x_k$ , y  $y_k$  de los rangos asignados a las variables.

El estadístico  $\rho$  viene dado por la expresión

$$\rho = 1 - \frac{6 \sum_{k=1}^n d_k^2}{n(n^2 - 1)}$$

La interpretación del coeficiente de correlación de Spearman es igual que la del coeficiente de correlación de Pearson. Oscila entre  $-1$  y  $+1$ , indicando asociaciones perfectas, inversas y directas respectivamente.

## 8.2. TABLAS DE CONTINGENCIA

Dadas las variables  $X$  y  $Y$  con  $m$  y  $s$  categorías mutuamente excluyentes, la tabla de doble entrada que las representa recibe el nombre de *tabla de contingencia de tamaño  $m \times s$* . El caso más sencillo de una tabla de contingencia es aquella de tamaño  $2 \times 2$ , es decir, cuando  $m = s = 2$  los valores de  $X$  e  $Y$  se mueven sobre dos categorías, este hecho hace que cada una de las variables de cuenta de la presencia o ausencia del atributo o característica estudiada.

Esta herramienta estadística es muy antigua y conocida, según lo indica Molinero (2004), es de gran utilidad en la comparación de dos o más grupos de sujetos con

respecto a una variable categórica, según Pita/Pértiga (2004), y también para estudiar el comportamiento de dos variables categóricas medidas en un grupo, permitiendo emitir juicios de asociación de variables principalmente cuando su naturaleza es de tipo cualitativa.

Gráficamente las tablas de contingencia de tamaño 2x2 se aprecian de la siguiente manera. Pita y Pértiga (2004).

**Tabla 3. TABLA DE CONTINGENCIA GENERAL PARA LA COMPARACIÓN DE DOS VARIABLES DICOTÓMICAS.**

VARIABLES		Presencia	Ausencia	Total
Y	X			
	Presencia		a	b
Ausencia		c	d	c + d
Total		a + c	b + d	n

En general, a partir de una tabla de contingencia  $m \times s$  se puede determinar la existencia o no de una relación entre las variables, en el caso que las variables estén asociadas estadísticamente la medida de la intensidad de dicha relación se calcula con base en los datos recolectados, mediante el coeficiente  $\chi^2$  que es apropiado para variables de tipo cualitativo.

### 8.3. INDEPENDENCIA ESTADÍSTICA

Apoyándose en Behar y Yepes (1996), se dirá que dos variables  $X$  y  $Y$  son *estadísticamente independientes*, si la distribución de la característica  $X$  es la misma en cualquier subconjunto de elementos definidos por la característica  $Y$ , en forma simétrica se podrá intercambiar  $X$  por  $Y$ .

Lo anterior puede escribirse con símbolos de varias formas:

$X$  y  $Y$  son estadísticamente independientes si:

$$h(x_i / y_j) = h(x_i) \text{ para todo } i, j \text{ con } 1 \leq i \leq m \text{ y } 1 \leq j \leq s.$$

lo cual implica que para cualquier  $x$ , se cumple:

$$h(x_i / y_1) = h(x_i / y_2) = \dots = h(x_i / y_s) = h(x_i)$$

De manera equivalente puede caracterizarse la independencia entre  $X$  y  $Y$ .

$$h(y_j / x_i) = h(y_j) \text{ para todo } i, j$$

Recordando la regla de la multiplicación que expresa:

$$h(x_i \cap y_j) = h(x_i) \times h(y_j / x_i)$$

Puede decirse que las variables  $X$  y  $Y$  son estadísticamente independientes si

$$h(x_i \cap y_j) = h(x_i) \times h(y_j), \text{ o lo que es igual } h_{ij} = h_{i*} \times h_{*j} \text{ para todo } i, j$$

Es decir cuando la distribución conjunta es el producto de las marginales, al encontrarse al menos una alguna pareja que no satisfaga la definición, es

suficiente para concluir que no hay independencia estadística, de lo contrario se concluirá que hay independencia.

Lo anterior es concluyente para determinar cuando hay o no independencia es decir si cumple o no la definición.

En la realidad existen grados o niveles de dependencia que deben ser medidos de manera tal que permita poner en evidencia la intensidad de la dependencia estadística. Puede pensarse en definir un instrumento que involucre la separación entre  $h(x_i \cap y_j)$  y el producto  $h(x_i) \times h(y_j)$  y que aumente el valor del indicador de dependencia, a medida que se separan los dos términos mencionados.

#### **8.4. CORRELACIÓN Y CAUSALIDAD**

Como lo expresa Runyon (1992), la correlación de dos variables permite predecir los valores de una variable basándose en los valores de la otra. Esta aplicación, a menudo conduce a un grave error de razonamiento; es decir, se piensa que la intensidad de la relación implica una dependencia funcional entre las variables. Lo que frecuentemente se pasa por alto es el hecho de que las variables pueden no estar unidas causalmente de una cierta manera, pero pueden variar juntas en virtud de un vínculo común con una tercera variable. Por ejemplo, se conoce que

existe una relación entre el peso y la talla de una persona, lo que no es cierto es que si la persona quiere aumentar su estatura lo logre aumentando su peso, de esta manera es claro que la correlación es una condición necesaria pero no suficiente para establecer una relación de causalidad entre dos variables. Esto mismo lo expresa Batanero (2001) de la siguiente manera: “*La existencia de correlación por sí sola no implica causación*”, en otras palabras se puede afirmar que la asociación estadística no implica necesariamente una relación causal, pero si ayuda a la búsqueda de la verdadera causa.

## **9. INCORPORACIÓN DE LA ESTADÍSTICA EN LA ESCUELA.**

Según Joshua (1993) las didácticas de las ciencias históricamente han nacido de la decisión de no dejar de tomar en consideración la especificidad de la disciplina. Estas se han fundamentado en una alternativa radical que debía distinguirlas progresivamente de otras aproximaciones referidas a la enseñanza de las ciencias: la voluntad y la afirmación de la posibilidad de una aproximación razonada, sistemática, científica y específica de los fenómenos de la enseñanza en estos campos, tendiente a delimitar teórica y prácticamente los campos de lo posible y aquellos de lo inaccesible.

Si se ha de arriesgar una definición, se podría decir que la didáctica de una disciplina es la ciencia que estudia, para un campo en particular, los fenómenos de enseñanza, las condiciones de la transmisión de la “cultura” propia de una institución y las condiciones de adquisición de conocimientos por parte de un aprendiz.

La didáctica de la estadística no ha sido ajena a este proceso de surgimiento y desarrollo, la estadística ha desempeñado un papel primordial en el desarrollo de la sociedad moderna, al proporcionar herramientas metodológicas generales para analizar la variabilidad, determinar relaciones entre variables, diseñar en forma

óptima estudios y experimentos y mejorar las predicciones y toma de decisiones en ambientes de incertidumbre. Existen además, diversas razones que justifican la importancia de la enseñanza de la estadística; las cuales han sido planteadas desde comienzos de la década de los 80, por diferentes autores, entre ellos Holmes (1980), quien señala:

- La estadística es una parte de la educación general deseable para los futuros ciudadanos adultos, quienes precisan adquirir la capacidad de lectura e interpretación de tablas y gráficos estadísticos que con frecuencia aparecen en los medios informativos.
- Es útil para la vida posterior, ya que en muchas profesiones se precisan unos conocimientos básicos del tema.
- Su estudio ayuda al desarrollo personal, fomentando un razonamiento crítico, basado en la valoración de la evidencia objetiva.
- Ayuda a comprender los restantes temas del currículo, tanto de la educación obligatoria, como posterior donde con frecuencia aparecen gráficos, resúmenes o conceptos estadísticos.

Batanero (1999) apoyada en Fischbein (1975) y Begg (1997) afirma que otro de los puntos por los cuales es importante la vinculación de la Estadística en la escuela, es la necesidad de mostrar al alumno una imagen más equilibrada de la

realidad, dado que el currículo de matemáticas tiene un carácter exclusivamente determinista. Otra razón es que la Estadística es un buen vehículo para alcanzar las capacidades de comunicación, tratamiento de la información, resolución de problemas, uso de ordenadores, trabajo cooperativo y en grupo, a los que hoy se da gran importancia.

Lo anterior resulta ser uno de los aspectos que ha mostrado la necesidad de la incorporación de la Estadística en los currículos escolares, y en particular en Colombia<sup>1</sup> dentro del área de Matemáticas se consignan cinco diferentes sistemas y pensamientos matemáticos, entre los cuales se encuentra, para el conocimiento estadístico, el denominado *pensamiento aleatorio y sistema de datos*, el cual busca desarrollar en los futuros ciudadanos las competencias básicas en esta área para generar una *cultura estadística*, que en términos de Batanero (2002) hace referencia a dos componentes interrelacionados que son:

- Capacidad para interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos estocásticos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos.

---

<sup>1</sup> País que no ha sido ajeno a este proceso ya que el Gobierno Nacional a través del Ministerio de Educación Nacional (MEN) ha diseñado los estándares de calidad para la educación en los niveles de formación de básica y media profesional

- Capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante. (Gal, 2002 en prensa, p.2-3).

Lo que se intenta dar a entender con lo referido en los anteriores componentes, es que no se debe buscar hacer del ciudadano un estadístico profesional, sino dotarlo de elementos básicos y necesarios, como ideas fundamentales sobre gráficos, resúmenes estadísticos, diseño de experimentos, diferencia entre estudios observacionales y experimentales, encuestas, incertidumbre y probabilidad y riesgo, entre otros.

Como respuesta a esta necesidad, los propios estadísticos, los psicólogos, y algunos investigadores en educación matemática se han interesado por investigar, desarrollar y proponer metodologías y procedimientos que permitan involucrar la Estadística dentro de la comunidad en general, desde la psicología, se conocen los estudios realizados acerca del razonamiento sobre la toma de decisiones en situaciones de incertidumbre y el desarrollo evolutivo del razonamiento estocástico. Dentro de la educación matemática, se ha investigado sobre la didáctica de la estadística, desde donde se han publicado algunos resultados de estas investigaciones.

Puesto que la estadística es una ciencia en constante cambio, es preciso que la enseñanza no se centre, únicamente en los contenidos específicos, sino que trate de desarrollar en los estudiantes una actitud favorable, un razonamiento

estadístico y un interés por complementar posteriormente su aprendizaje. Para alcanzar este fin en la enseñanza de la estadística, los investigadores en educación estadística han planteado:

- Que los alumnos lleguen a comprender y a apreciar el papel de la estadística en la sociedad, conociendo sus diferentes campos de aplicación y el modo en que la Estadística a contribuido a su desarrollo.
- Que los alumnos lleguen a comprender y a valorar el método estadístico, esto es, la clase de preguntas que un uso inteligente de la Estadística puede responder, las formas básicas de razonamiento estadístico, su potencia y limitaciones.

Esta metodología de enseñanza consistente en generar desde la escuela una cultura estadística, se soporta en las teorías de aprendizaje que enfatizan el papel de la resolución de problemas, de la actividad del alumno en la construcción del conocimiento, así como de la formulación (lenguaje matemático), validación (demostración y razonamiento de la ideas matemáticas) e institucionalización (puesta en común; acuerdo social en la construcción del conocimiento). Batanero (2001 a) expresa que en esta metodología el profesor ya no es un trasmisor del conocimiento, sino un gestor de este conocimiento y del medio (instrumentos, situaciones) que permita al alumno progresar en su aprendizaje.

Los proyectos estadísticos y la experimentación con fenómenos aleatorios cobran un papel primordial, puesto que estos proyectos permiten a los alumnos elegir un tema de interés en el cual precisan definir los objetivos, elegir los instrumentos de recolección de datos y darles un tratamiento estadístico. Además los proyectos introducen a los alumnos en la investigación, les permite apreciar la dificultad e importancia del trabajo estadístico y les hace interesarse por la estadística como medio para abordar problemas variados de la vida real.

La vinculación de la Estadística en el currículo escolar, ha despertado el interés de diferentes instituciones como el *International Statistical Institute* (ISI) y su brazo educativo el *International Association for Statistical Education* (IASE) quien tiene por objetivo principal el desarrollo y difusión de recursos para la enseñanza de la Estadística; desde esta perspectiva conjuntamente con la propia Estadística surge lo que se conoce como *Educación Estadística*. Al respecto Batanero (2001 a) la define como “El campo de innovación, desarrollo e investigación constituido por todas aquellas personas (educadores estadísticos) que se interesan o trabajan por mejorar la enseñanza, el aprendizaje, la comprensión, la valoración, el uso y las actitudes hacia la Estadística”. Esta definición plasma los intereses y esfuerzos realizados por lograr que la Educación Estadística se consolide como una disciplina.

Una de las organizaciones internacionales más antiguas en Estadística es el *ISI* (*International Statistical Institute*). La cual se constituyó en 1885 como una sociedad autónoma que intenta desarrollar y mejorar los métodos estadísticos y su

uso. Hoy en día, el ISI está constituido por 7 secciones: International Association of Survey Statistician (IASS), Bernoulli Society for Mathematical Statistics and Probability, International Association for Statistical Computing (IASC), International Association for Official Statistics (IAOS), Irving Fisher Society on Financial and Monetary Statistics, International Society for Business and Industrial Statistics (ISBIS), y el International Association for Statistical Education (IASE), creado en 1991 como evolución del Comité de Educación fundado en 1948 por el ISI, quien en este tiempo preocupado por la situación del mundo a causa de la posguerra, creó el Comité con el objetivo de formar estadísticos profesionales que ayudaran con la economía y desarrollo de los países afectados, pero en este proceso se evidenciaron dificultades en la enseñanza y el aprendizaje de la estadística, entonces decide crear el IASE cuyo objetivo principal es contribuir al desarrollo y mejora de la Educación Estadística en el ámbito Internacional. En la actualidad el IASE cuenta con unos 500 miembros, que son personas interesadas en la enseñanza de la estadística en al menos uno de los niveles educativos.

Existen otras sociedades estadísticas establecidas y funcionando en la actualidad a nivel mundial, pero con carácter más regional, como por ejemplo la American Statistical Association (ASA) fundada en Estados Unidos en 1839, y que igualmente ha mostrado preocupación por la evolución de la Estadística. El IASE cobra importancia dado que esta institución es la que involucra una reflexión teórica de la educación estadística e intenta llevarla a los países menos desarrollados del planeta, a través de la creación, el desarrollo, la difusión y la convocatoria a

diversos eventos que apuntan hacia la construcción de conocimiento alrededor de la estadística y los procesos de enseñanza y aprendizaje de esta.

El IASE, dentro de sus mecanismos de difusión viene realizando diferentes actividades entre las que se encuentran las conferencias, en particular la International Conference of Teaching Statistical (ICOTS), donde se han discutido temas como “Educación en la escuela”, “Educación postsecundaria”, y “Perspectiva internacional de la Educación Estadística”, entre otras. Las ICOTS a nivel mundial convocan a toda la comunidad de educadores estadísticos para que presenten proyectos y resultados de investigación, al tiempo que se discute el papel fundamental que cumple la estadística actualmente en la sociedad.

El interés de este seminario radicó en conocer problemáticas que se han generado en el estudio del proceso de construcción del significado del objeto estadístico *asociación estadística*. Para lo cual, inicialmente, se apropió un marco teórico de la Educación Matemática, debido a que aún no existen teorías establecidas en didáctica de la estadística ya que ésta es un área en formación y los marcos y metodologías que se están implementando en este campo tienen su origen en la didáctica de la matemática, que es el área mas próxima.

Godino (1994), en su propuesta teórica: “Significado de los Objetos Matemáticos” define la Didáctica de la Matemática como “La disciplina encargada de estudiar los procesos de enseñanza/aprendizaje de los saberes matemáticos, en los aspectos teórico-conceptuales y de resolución de problemas, tratando de caracterizar los

factores que condicionan dichos procesos. Se interesa por determinar el significado que los alumnos atribuyen a los términos y símbolos matemáticos, a los conceptos y proposiciones, así como la construcción de estos significados como consecuencia de la instrucción”.

Este seminario está enmarcado en dos áreas: la Estadística y la Didáctica de la Estadística, y en esta última, principalmente está enfocado en lo que respecta al aprendizaje, y específicamente en la comprensión del objeto *asociación estadística*.

## 10. UNA PROPUESTA TEÓRICA PARA EL ANÁLISIS DIDÁCTICO

Godino y Batanero (1994), ofrecen una propuesta para entender como se construye el significado de los objetos matemáticos. Proponen un triple condicionamiento para la matemática: como actividad de resolución de problemas socialmente compartida, como lenguaje simbólico y como sistema conceptual lógicamente organizado. Ellos parten esta teorización de *la noción de problema* formulada por Lester (1980) como: “una situación en la que se le pide a un individuo realizar una tarea para la que no tiene un algoritmo fácilmente accesible que determine completamente el método de solución”, continuando con lo que se entiende por *campo de problemas* como “agrupación de problemas matemáticos que comparten conceptos y modos similares de resolución”. En la resolución de estos problemas se precisa *matematizar*, es decir “buscar dentro del campo de problemas lo esencial entre los distintos contextos, las relaciones con otras situaciones, problemas o procedimientos; es también propio de esta actividad comunicar a otros la solución, validarla y generalizarla a otros contextos y problemas”. Cuando una persona se enfrenta a la resolución de un problema realiza una *Práctica personal* “que es toda actuación o manifestación, lingüística o no, realizada por alguien para resolver un problema matemático” y se llama *práctica personal significativa* “cuando esta actividad tiene sentido para el individuo”. Se habla o entiende por *Institución* a “un grupo de personas que se

dedica a realizar actividades para resolver problemáticas del área de la matemática que comparte con otras ciencias de acuerdo con unos modos de resolución también compartidos”. *Las prácticas institucionales*, “se dan cuando estas son aceptadas por la institución dada su pertinencia”. *Los Objetos matemáticos institucionales* son “entes abstractos que surgen de las prácticas institucionales”. *Un sistema de prácticas personales* “es la serie de prácticas significativas que realiza una persona cuando resuelve un campo de problemas”. *Los objetos personales* “emergen de los sistemas de prácticas personales asociadas a un campo de problemas”. Se denomina *significado personal-institucional del objeto* “a los sistemas de prácticas asociados a un campo de problemas de donde emergen los objetos personales-institucionales”.

Los *sistemas de prácticas* se entienden como “los elementos del objeto que se ponen en juego en la actividad matemática” los cuales son elementos del significado (enunciados, notaciones, símbolos, operaciones, conceptos, proposiciones, demostraciones, comprobaciones y otros), en consecuencia cada elemento del significado se caracteriza por un sistemas de prácticas.

De la comparación del significado del objeto personal con el institucional se deduce el estado del conocimiento del sujeto sobre un objeto matemático.

Lo anterior constituye una propuesta para el Análisis Didáctico, dado que en esta teorización Juan Díaz Godino define formalmente lo que se entiende por significado de un objeto matemático y establece los elementos constitutivos de

éste, los cuales permiten evidenciar la comprensión del significado del objeto a partir de la caracterización de estos elementos.

## 11. LA ASOCIACIÓN ESTADÍSTICA DESDE LA PERSPECTIVA DIDÁCTICA

Estepa y Sánchez (2000) basándose en los estudios realizados por Inhelder y Piaget (1955), apuntan que el concepto de *asociación estadística*, no se construye intuitivamente, ya que este es un concepto que no tiene ligado a él un referente tangible a pesar de ser un concepto fundamental y considerado como básico en Estadística, su comprensión requiere además de su definición matemática, de un campo de problemas y de un sistema de prácticas significativas ligadas a éste.

Según Estepa y Sánchez (2000), Hildebrand y cols (1977), afirman que “Uno de los conceptos estadísticos de gran importancia tanto en si mismo, como por ser prerrequisito para continuar avanzando en el conocimiento estadístico es la *Asociación Estadística*, entendiendo por tal la extensión del concepto de correlación a variables cualesquiera, incluso no numéricas”.

En el sentido de Godino (1996), el significado del objeto *asociación estadística*, es el sistema de practicas institucionales asociadas al campo de problemas de las que emerge la asociación estadística, que para este objeto es muy extenso, en general son las ciencias, y dentro de cada una de ellas se puede encontrar un campo de problemas especifico, que para los intereses de este seminario se ha

tomado de dos de las ciencias interesadas en la cognición humana: la Psicología y la Didáctica de la Estadística.

Según Godino (1996) para apropiarse el significado institucional del objeto *asociación estadística* se hace necesario asir los siguientes elementos extensionales, intencionales e instrumentales, constituyentes del significado de los objetos matemáticos, entendidos como:

- *Extensionales*: Los diferentes problemas y situaciones prototípicas (situación modelo) donde se usa el objeto, en este caso son: tablas de contingencia, diferencia de medias, y gráficos de dispersión.
- *Intencionales*: las diferentes propiedades, características y relaciones de los objetos matemáticos con otras entidades, para el objeto asociación se encuentran: coeficientes de correlación, medidas de tendencia central, funciones, independencia de variables, probabilidades, probabilidades condicionales, proporción, combinatoria, regresión lineal, entre otras.
- *Instrumentales*: las diferentes herramientas semióticas (símbolos) disponibles para estudiar, resolver y/o representar los problemas y los objetos matemáticos involucrados. En este caso se encuentran: tablas de distribución conjunta de frecuencias absolutas y relativas, tablas de distribuciones marginales de frecuencia, líneas de regresión, gráficos de funciones, tablas de contingencia.

Como lo expresa Estepa (2000) a pesar de la importancia estadística y curricular del objeto asociación estadística, la investigación llevada a cabo, de interés para la enseñanza del tema, es escasa y procede fundamentalmente de la Psicología y de la Didáctica de la Matemática.

En cuanto a Educación Estadística, se han desarrollado estudios desde la Psicología, la Estadística y la Educación Matemática, autores como Carmen Batanero, Antonio Estepa, Francisco Sánchez y otros se han preocupado por las cuestiones concernientes a los procesos de enseñanza y aprendizaje de la Estadística, en conceptos básicos, y en particular lo referente a la *asociación estadística*.

A continuación se presentan los resultados de algunas investigaciones que relacionan el aprendizaje del objeto *asociación estadística*, algunas de ellas realizan su análisis a la luz del marco de Godino.

### **11.1. INVESTIGACIONES DIDÁCTICAS DESDE LA PERSPECTIVA PSICOLÓGICA**

Desde el campo de la Psicología los estudios realizados sobre la comprensión del objeto *asociación estadística* fueron iniciados por Inelhder y Piaget (1955), autores referenciados por Estepa (2000), consideran importante la investigación en lo referente al desarrollo evolutivo de este objeto por considerar que tiene conexiones con el pensamiento causal, además ponen de manifiesto que apropiar

el concepto de *asociación estadística* es el último paso para comprender el concepto de probabilidad. En consecuencia, los antecedentes evolutivos de los dos conceptos son los mismos y la comprensión de la idea de asociación implica para estos autores la de proporción y probabilidad, requiriendo la capacidad combinatoria.

Emitir juicios acerca de la *asociación estadística* es una actividad importante para la toma de decisiones en diversos campos. En la psicología, el estudio sobre este tipo de juicios se ha centrado en las tablas de contingencia 2x2 como campo de problemas, que es el caso más simple de asociación estadística que se puede definir, ya que las dos variables cualitativas consideradas son dicotómicas. Estas investigaciones, según Estepa y Sánchez (2000), se pueden clasificar, dependiendo del objeto del estudio, en: a) Estrategias empleadas para resolver el problema. Por ejemplo, Pérez Echeverría (1990) estudia siete tipos de estrategias, según las celdas de las tablas de contingencia utilizadas para resolver el problema, agrupándolas en cinco niveles, llegando a la conclusión que las estrategias de tipo normativo eran poco utilizadas; b) influencia de las teorías previas, en estas investigaciones se observa como los sujetos, al realizar el juicio de asociación, se guían más por sus creencias sobre el contexto del problema que por los datos presentados en el problema. Se consideran dos conceptos relacionados con las teorías previas: correlación ilusoria “*cuando percibimos que existe una correlación basándonos en nuestras propias teorías, pero no existe ningún hecho empírico que las sustente*”. Murphy y Medin, (1985) e ilusión de control “*como ilusión de control se define la expectativa inapropiada de que la*

*probabilidad de un éxito personal sea mas alta que lo que garantiza la probabilidad objetiva*" (Langer,1975); c) exactitud de los juicios de asociación, en general, no suelen ser muy exactos dependiendo de varios factores como signo e intensidad de la asociación, datos ofrecidos y forma de presentarlos y d) el contexto y la presentación de la información también influyen en los juicios de asociación realizados.

## **11.2. INVESTIGACIONES DIDÁCTICAS A PARTIR DE LA DIDÁCTICA DE LA MATEMÁTICA**

Como se ha indicado con anterioridad a pesar de la importancia del objeto *asociación estadística* tanto a nivel disciplinar, como de sus aplicaciones, los pioneros en investigación en Didáctica de la Estadística señalan que es poca la investigación en este tema, pero uno de los campos donde es posible encontrar literatura al respecto es en Didáctica de la Matemática, que por ser una disciplina con suficiente formación cuenta con investigaciones alrededor del origen y desarrollo de los conceptos matemáticos y es justamente desde allí donde los estudiosos han concluido la existencia de una primera problemática epistemológica propia del objeto asociación estadística debido a su complejidad y que como se ha indicado en estudios psicológicos y didácticos es un concepto que necesita de escolaridad para su comprensión.

Batanero y otros (1998) presentan una de las problemáticas que surge en la apropiación de los elementos del significado institucional del objeto, utilizando

ordenadores, se llevan a cabo experimentos de enseñanza poniendo de manifiesto la dificultad de la tarea debido a su complejidad.

La confrontación de este planteamiento se hace a través de dos experimentos bietápicos con dos grupos de estudiantes de primeros cursos universitarios que asisten al curso de estadística, uno en el nivel descriptivo y otro en el nivel inferencial. En el primer grupo la primera fase consiste en detectar los preconceptos de asociación a través de situaciones problema y destacar las concepciones erróneas a la luz de los elementos extensionales e intencionales del significado, tales como:

Concepción determinista de la asociación: Cuando el estudiante solo admite para un valor de la variable independiente, un solo valor para la variable dependiente, algunos estudiantes al observar que existe mas de un valor para la variable independiente, declaran la independencia entre las variables. Por ejemplo el estudiante no admite que existe una relación entre la talla y el peso de una persona, dado que para determinar la talla existen varios valores del peso (subir o bajar de peso en un intervalo de tiempo inclusive durante un mismo día).

Concepción unidireccional de la asociación: el estudiante percibe la existencia de asociación solamente cuando es positiva, considerando la negativa como independencia. Por ejemplo, el alumno no evidenciaría que existe relación entre los niveles de ingreso de los vecindarios y sus tasas de delitos (a *mayor* ingreso, *menor* tasa de delitos).

Concepción local de la asociación: cuando el estudiante realiza el juicio de asociación basándose en parte de los datos presentados y no en todos; si la parte de datos considerada presenta un tipo de asociación, adopta este tipo para todo el conjunto de datos. Se ilustra esta situación con el siguiente ejemplo: Se desea conocer si un medicamento produce trastornos digestivos. Para ello se ha observado durante un periodo de tiempo a una muestra de 250 individuos, los resultados se consignan en la siguiente tabla:

Trastornos Droga	TIENE	NO TIENE	Total
TOMA	90	60	150
NO TOMA	60	40	100
Total	150	100	250

El estudiante considera que existe relación entre tomar un medicamento y padecer trastornos digestivo basado en sólo una porción de los datos que corresponde a la celda de mayor frecuencia.

Concepción causal de la asociación: cuando el estudiante exige la existencia de relación causal para afirmar la existencia de asociación. Por ejemplo, existe una asociación entre la talla y el peso de una persona, sin embargo esta relación no es causal, es decir si una persona desea aumentar su estatura no lo logra aumentando su peso, o perder peso no se logra perdiendo estatura, hecho ante el cual el estudiante decide declarar independencia entre las variables.

En la segunda fase después de utilizar los ordenadores como herramienta facilitadora (elementos instrumentales) se valora el nivel de comprensión de los elementos mediante problemas para evidenciar el aprendizaje del concepto de asociación a través de la apropiación de los tres elementos del significado antes mencionados.

En el segundo grupo se realizan las mismas dos fases de manera similar aunque los problemas ahora tienen que ver con objetos de la inferencia, (intervalos de confianza, test de chi cuadrado, diferencia de medias) y destacan en las conclusiones que después de la enseñanza persisten las concepciones erróneas, tales como la concepción causal.

## 12. RESULTADOS DE LAS PROBLEMÁTICAS

Entendiendo el concepto de problemática como el conjunto de problemas pertenecientes a una ciencia o actividad determinada, (R.A.E. 1992), para efectos de este trabajo se considera como problemática el conjunto de problemas ligados a la construcción del significado del objeto estadístico “asociación de variables”.

Como resultado de este seminario se describen las problemáticas que se han abordado en las investigaciones tituladas “Construcción del Significado de la Asociación Estadística Mediante Actividades de Análisis de Datos” (Batanero 1998) y “Correlación y Regresión en los Primeros Cursos Universitarios” (Estepa y Sánchez 2000) basadas en el marco teórico sobre el significado y comprensión de los objetos matemáticos que se describen en Godino (1996) y Godino y Batanero (1994, 1998), las que refieren e identifican los conceptos relacionados en la teorización de Godino desde las investigaciones ya mencionadas, por último, se analizan los resultados obtenidos en estas.

## 12.1. CONSTRUCCIÓN DEL SIGNIFICADO DE LA ASOCIACIÓN ESTADÍSTICA MEDIANTE EL ANÁLISIS DE DATOS

Los conceptos que se tratan en este documento son:

*La institución:* desde el punto de vista de personas interesadas en resolver un campo de problemas, en este experimento se identifica como el curso universitario introductorio al análisis de datos y estadística descriptiva.

*Elementos del Significado del objeto “Asociación Estadística”*

En esta investigación se identifican los siguientes *elementos extensionales o campos de problemas:*

- Realizar un juicio de asociación en una tabla de contingencia general para la comparación de dos variables dicotómicas.

**Tabla No. 4. DEFINICIÓN CLÁSICA DE UNA TABLA DE CONTINGENCIAS 2X2**

VARIABLES	Presencia	Ausencia	Total
$X$			
$Y$			
Presencia	a	b	a + b
Ausencia	c	d	c + d
Total	a + c	b + d	N

- Valorar la correlación existente entre dos variables de la misma naturaleza y dos variables de naturaleza distinta.

- Averiguar si una variable numérica tiene la misma distribución en dos muestras diferentes.

**Tabla No. 5. TABLA DE DISTRIBUCIÓN DE UNA VARIABLE PARA DOS MUESTRAS**

Sujetos grupo	$S_1$	$S_2$	$S_3$	...	$S_i$	...	$S_n$
Grupo 1	$V_{11}$	$V_{12}$	$V_{13}$	...	$V_{1i}$	...	$V_{1n}$
Grupo 2	$V_{21}$	$V_{22}$	$V_{23}$	...	$V_{2i}$	...	$V_{2n}$

Donde

$S_i$ : representa el sujeto  $i$  a tratar

$V_{ij}$ : representa el valor del sujeto  $i$  en el grupo  $j$

Dentro de *los elementos intencionales* que se incluyen en los ítems propuestos para el experimento que realizaron los grupos se encuentran: conceptos básicos sobre poblaciones y muestras, organización de datos, tipos de variables estadísticas y distribuciones de frecuencias, gráficos, parámetros y estadísticos, variables estadísticas bidimensionales, tablas de contingencia, covarianza, correlación y regresión lineal, conceptos de muestreo, distribución en el muestreo, intervalos de confianza, test de hipótesis de las medias para una y dos muestras y test de chi-cuadrado.

Los *elementos instrumentales* o representacionales, utilizados en este experimento son las herramientas del paquete estadístico statgraphics, que se

pueden clasificar por su nivel de reducción de datos, su carácter numérico o gráfico y su enfoque analítico (descriptivo e inferencial)

### *Prácticas o estrategias*

Las prácticas personales significativas, que utilizan los sujetos para solucionar los problemas propuestos son:

- Utilizar la tendencia constante, creciente o decreciente de los puntos en los diagramas de dispersión para justificar el tipo de asociación (nula, positiva o negativa).

### **Representación gráfica de una relación**



- Utilizar las medias o los totales para comparar la distribución de una variable en dos muestras diferentes.
- Comparar las frecuencias de casos a favor y en contra de la asociación en cada valor de la variable independiente o la razón de estas frecuencias en tablas de contingencia 2x2.

Se puede establecer un paralelo entre las prácticas personales y las practicas institucionales en el proceso de apropiación de algunos elementos intencionales del significado de la asociación.

**Tabla No. 6. COMPARACIÓN DE LAS PRÁCTICAS INSTITUCIONALES Y LAS PRÁCTICA PERSONALES.**

PRACTICAS INSTITUCIONALES	PRACTICAS PERSONALES
La comparación de dos o más muestras, con objeto de estudiar la posible relación entre dos variables debe efectuarse en términos de frecuencias relativas	Los estudiantes inician comparando frecuencias absolutas de la distribución de una variable en dos muestras.
La posible existencia o no de diferencias en la distribución de una variable entre dos o mas muestras se deduce a partir de la comparación de toda la distribución de la variable en cada una de las muestras y no de una parte de la misma.	Los estudiantes comparan valores aislados, al estudiar las dos muestras.
A partir de una misma frecuencia absoluta pueden deducirse dos frecuencias relativas condicionales diferentes, según la variable que se emplee como condición. El papel de condición y condicionado en la frecuencia relativa condicional no es intercambiable.	Los estudiantes no diferencian en ocasiones, el papel que desempeña la condición y el condicionado, es decir confunden la probabilidad de A dado B, $P(A/B)$ , con la probabilidad de B dado A, $P(B/A)$ , o no las discriminan.
Dos variables son independientes, si la distribución condicional de una de ellas no cambia cuando se varían los valores de la otra variable	Los estudiantes no descubren que una condición para la independencia es la invarianza de las distribuciones marginales, cuando varía el valor de la variable condicionante.
En la determinación de la asociación entre dos variables, estas desempeñan un papel simétrico. Por el contrario, en el estudio de la regresión las variables desempeñan un papel asimétrico. Existen dos rectas de regresión diferentes, según cual de las dos variables actúe como predicción y de respuesta	No distinguen la simetría y la asimetría de asociación y la regresión donde se pueden encontrar dos líneas para las mismas dos variables, pero el coeficiente de asociación (correlación) es el mismo.
Una correlación positiva, indica dependencia directa entre las variables	Interpretan la magnitud (intensidad) cercana a 1, pero no identifican la relación de monotonía entre las variables (ambas crecen, ambas decrecen)
Una correlación negativa indica dependencia inversa entre las variables	No esperan coeficientes negativos, no comparan los coeficientes negativos (obstáculo epistemológico) se tiende a pensar que a menor magnitud mayor asociación (orden en los enteros)
El valor absoluto del coeficiente de correlación es indicativo de la intensidad de la asociación	Aunque se asocia la magnitud a la intensidad de la relación, no la identifican con el coeficiente.

### *Problemáticas de tipo Didáctico*

Desde la perspectiva de las problemáticas de tipo didáctico, los autores al realizar los experimentos identifican las concepciones erróneas sobre la asociación estadística que revelan los estudiantes en las estrategias utilizadas al resolver el campo de problemas.

Concepción determinista de la asociación: tienen un concepto funcional desde el punto de vista matemático para la asociación, es decir no se admite más de un valor de la variable dependiente para cada valor de la variable independiente.

Concepción unidireccional de la asociación: cuando solo se considera la existencia de la relación, en los casos que la relación es concordante, considerando la discordante como independencia estadística.

Concepción local de la asociación: cuando se infiere la existencia o ausencia de la relación dependiendo de la mayor frecuencia absoluta.

Concepción causal de la asociación: cuando se acepta la relación entre dos variables si un cambio en la variable de predicción implica un cambio en la variable de respuesta.

### *Problemática de tipo Psicológico*

De acuerdo con la tabla No. 4, los investigadores sobre la comprensión del concepto de asociación estadística, en el campo de la psicología, Inhelder y Piaget encontraron que los estudiantes que tienen un conocimiento previo en el cálculo de probabilidades sencillas al enfrentarse a una tabla de contingencia para emitir un juicio de asociación entre variables dicotómicas, solo tenían en cuenta el caso para el cual los valores de ambas variables son favorables al juicio de asociación demandado (celda a), en otros casos solo comparaban las celdas dos a dos, este hecho se explica observando que la comprensión de la asociación requiere considerar los casos favorables, es decir cuando las relaciones entre las variables es directa, positiva o concordantes (celdas a y d) y las relaciones indirectas, negativos o discordantes (celdas b y c).

## **12.2. CORRELACIÓN Y REGRESIÓN EN LOS PRIMEROS CURSOS UNIVERSITARIOS**

Para esta investigación los conceptos que se identifican son:

*La institución:* esta conformada por estudiantes de primeros cursos universitarios que se están preparando en estadística y pertenecen a diferentes programas.

*Elementos del Significado del objeto “Asociación Estadística”*

Los elementos extensionales o campo de problemas para esta investigación son:

- Calcular la media de la variable dependiente a partir de la media de la variable independiente, conociendo la recta de regresión. (En el problema no se precisa cual variable se considera como independiente y cual como dependiente )

En una tabla como la siguiente, presentan la distribución de los valores de dos variables, se pide al estudiante:

**TABLA No. 7 DISTRIBUCIÓN PARA LOS VALORES DE DOS VARIABLES**

Variable $X$	$x_1$	$x_2$	$x_3$	...	$x_i$	...	$x_n$
Variable $Y$	$y_1$	$y_2$	$y_3$	...	$y_i$	...	$y_n$

- Calcular el coeficiente de correlación lineal.
- Determinar el tipo de relación (Directa, inversa, o independenciam) que existe entre las variables.
- Determinar la recta de regresión de  $Y$  sobre  $X$  .
- Se pide predecir un valor  $y_i$  de  $Y$  , dado el valor  $x_i$  de  $X$  , teniendo que  $x_1 \leq x_i \leq x_n$  , y también en el caso para el cual  $x^* \geq x_n$  .
- Se pide predecir un valor  $x_i$  de  $X$  , dado el valor  $y_i$  de,  $Y$  .

Como *elementos intencionales* se consideran los contenidos del programa de estadística que estudiaron los alumnos que conformaron la muestra para el

experimento, los cuales son: tablas de frecuencias y gráficos, medidas de centralización y dispersión, simetría y curtosis, variables estadísticas bidimensionales, tablas de contingencia, covarianza, correlación, regresión lineal y polinómica, muestreo, distribución en el muestreo, intervalos de confianza y test de hipótesis.

Los *elementos instrumentales* los constituyen: La calculadora, el diagrama de dispersión, el algoritmo del coeficiente de correlación, las expresiones matemáticas de las rectas de regresión.

### *Prácticas o estrategias*

Entre las practicas personales realizadas se encuentran:

- Calculan la media de la variable dependiente usando la recta de regresión, se distinguen tres casos:
  - La expresión explícita de la recta de regresión ( $Y$  sobre  $X$  y  $X$  sobre  $Y$  ).
  - Usan la expresión punto pendiente de la recta de regresión.
  - Usan combinaciones de las expresiones de las rectas de regresión y las expresiones punto pendiente de las mismas.

- Calculan el coeficiente de correlación usando el algoritmo correcto, aunque las respuestas en algunos casos sean incorrectas dado el desconocimiento del uso apropiado de la calculadora.
- Para determinar la recta de regresión de  $y$  sobre  $x$  los estudiantes utilizan las recta de regresión de  $X$  sobre  $Y$ ,  $Y$  sobre  $X$  o ambas.
- Para determinar el tipo de relación existente entre las variables los estudiantes utilizan estrategias como:
  - Usan el coeficiente de correlación
  - Variación conjunta de las variables (monotonía)
  - Uso de la covarianza
  - Usa el coeficiente de determinación, (el modelo ajustado a los datos)
  - Usan el diagrama de dispersión, un intervalo de variación del coeficiente de correlación o comparan la asociación y la proporcionalidad.
- En la predicción de un valor para la variable de respuesta, dado un valor de predicción, tanto en el intervalo de variación como por fuera de este, los estudiantes hacen uso de las siguientes estrategias:
  - Recta de regresión de  $Y$  sobre  $X$ .
  - Recta de regresión de  $X$  sobre  $Y$
  - El concepto de proporcionalidad

### *Problemática de tipo Didáctico*

A continuación, se exponen los errores y dificultades que presentan los estudiantes al intentar resolver el campo de problemas propuesto en esta investigación, se considera que el conjunto de estas dificultades constituye una problemática de tipo didáctico:

- Para determinar el coeficiente de correlación a partir de una tabla de datos, algunos alumnos dibujan un diagrama de dispersión para intuir el resultado pedido. Es interesante especificar el siguiente razonamiento: “el resultado no es correcto, porque no está comprendido en el intervalo  $[-1,1]$ ”, pero no es capaz de calcular el correcto; a la vista del diagrama de dispersión.
- Al calcular la media de la variable dependiente, los estudiantes utilizan las dos rectas de regresión ( $X$  sobre  $Y$  y  $Y$  sobre  $X$ ), la dificultad más notable que se presenta en los estudiantes al intentar resolver este problema es confundir la variable dependiente con la independiente.

La ecuación general de la recta de regresión de  $Y$  sobre  $X$  está dada por:

$$y = mx + n$$

- Al recurrir a esta expresión para enfrentarse a los problemas del cuestionario aparece la dificultad de la incorrecta interpretación de los parámetros  $m$  y  $n$  de la ecuación de la recta.

- Algunos estudiantes manifiestan confusión entre la intensidad y el sentido de la correlación entre las variables.
- Los alumnos al resolver problemas que implican hacer predicciones, se interesan más en realizar el cálculo y obtener una respuesta, más que en darle sentido al resultado en el contexto.
- Los estudiantes son escasamente conscientes de las restricciones que deben imponerse a la extrapolación.

#### *Problemática de tipo Psicológico*

- Dentro de las investigaciones psicológicas utilizadas como marco de referencia para la realización de este experimento se encuentra la de Inhelder y Piaget (1955), descrita anteriormente.

### 13. CONCLUSIONES

Después de revisar la literatura referente a lo que se entiende por asociación estadística, como se construye su significado y reconocer problemáticas que se presentan en el proceso de apropiación de este concepto, se concluye que:

- El objeto asociación estadística es considerado un concepto de gran importancia dentro de la estadística, ya que tiene conexiones con otras áreas del conocimiento
- Por ser un concepto fundamental es de gran utilidad para continuar avanzando en el conocimiento estadístico.
- La investigación en Didáctica de la Estadística es escasa, lo que supone que al nivel del objeto asociación estadística es aun menor dada la especificidad del concepto, y que los trabajos que se registran proceden básicamente de la psicología, la estadística y la didáctica de la matemática.
- Desde el ámbito escolar las problemáticas que son reconocidas por los investigadores en Didáctica de la estadística son de tipo epistemológico, psicológico y didáctico.

- A partir de la Didáctica de la matemática se ha elaborado una teoría alrededor del significado de los objetos matemáticos, en la cual se han soportado algunos experimentos sobre la construcción del significado de asociación estadística.
- La teorización sobre el significado de los objetos matemáticos, establece los elementos que constituyen el significado del objeto y son perceptibles, los cuales permiten evidenciar el estado del conocimiento del sujeto sobre el objeto matemático.
- La falta de reconocimiento del estadístico en los grupos de investigación, aunado a la falta de personal calificado que oriente el área, constituyen otras problemáticas en los procesos de enseñanza aprendizaje.
- Al abordar la resolución de una problemática, no existen propuestas únicas y definitivas, sino una diversidad de alternativas acordes a la especificidad del área.
- En la literatura revisada, se encuentra aplicado el marco teórico sobre comprensión del significado de los objetos matemáticos a investigaciones diseñadas para primeros cursos universitarios, mientras que en educación básica y media no se reportan resultados.

## 14 IBLIOGRAFÍA

AZCÁRATE Godeo Pilar, Cardeñoso, Domingo José María. La Naturaleza de la Matemática escolar: problema fundamental de la Didáctica de la Matemática. Departamento de Didáctica. Universidad de Cádiz y Universidad de Granada. España. En Investigación en la Escuela No. 24, 1994 p. 79-88.

BATANERO Carmen, Godino Juan Díaz, Green D.R. Holmes, P Vallecillos, Errores y Dificultades en la Comprensión de los Conceptos Estadísticos Elementales. International Journal of Mathematics Education in Science and Technology 1994.

BATANERO Carmen, Godino Juan Díaz y Estepa Antonio. Construcción del Significado de la Asociación Estadística Mediante Actividades de Análisis de Datos, Departamento de Didáctica de la Matemática, Universidad de Granada, Universidad de Jaén Madrid España. 1998

BATANERO, Carmen. ¿HACIA DÓNDE VA LA EDUCACIÓN ESTADÍSTICA?, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España. 1999.

BATANERO Carmen, Didáctica de la Estadística, Grupo de Educación Estadística, Departamento de Didáctica de la Matemática Universidad de Granada Madrid España, 2001. p. 29-33, 100-106

BATANERO Carmen. Presente y Futuro de la Educación Estadística, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España 2001A.

BATANERO, Carmen. Los Retos de la Cultura Estadística, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España 2002.

BEHAR G Roberto y Yepes A. Mario. Estadística Un Enfoque Descriptivo, Universidad de Valle, Facultad de Ingeniería, Departamento de Producción e Investigación de Operaciones,. Talleres Gráficos de Impresora, Fenusa S. A. Cali Colombia, 1996. Capítulos III, IV. p. 87-151, 157-209

CANAVOS George C. Probabilidad y Estadística Aplicación y Métodos, Virginia Commonwealth University, McGraw-Hill / Interamericana de México S.A. De C. V. 1988. Capítulo XIII p. 443 – 571.

CASTILLO M Alberto y Ojeda R. Mario Miguel, Principios de Estadística no Paramétrica, 1ra edición, Universidad Veracruzana, Xalapa, Veracruz., Mexico , octubre de 1994. p.143-149.

CAZU Pablo. La Investigación Bibliográfica, Buenos Aires Argentina, marzo 2000.

ESTEPA Antonio y Sánchez Francisco. Correlación y Regresión en los Primeros Cursos Universitarios. Departamento de Didáctica de las Ciencias, Departamento de Matemáticas, Escuela Politécnica Superior. Universidad de Jaén. Madrid. España. 2000

FERRIS J. Ritchey, Estadística para las Ciencias Sociales, “El Potencial de la Imaginación Estadística”, Mc Graw Hill, 2002 p.365-370

GODINO, Juan Diaz y Batanero, Carmen. Significado Institucional y Personal de los Objetos Matemáticos, Recherches en Didactique des Mathématiques, Vol 14 Revista No. 3 p. 325 – 355. 1994

JOSHUA Samuel, Jacques Jean Dupin 1993, Introducción a la didáctica de las ciencias y las matemáticas, PUF, Paris, p 1-10. Traducción y Adaptación Gloria Castrillón Castro y Miriam Vega Restrepo, Grupo de Educación Matemática, Instituto de Educación y Pedagogía , Universidad del valle Cali Colombia 1998

MOLINERO Luis M. Asociación de Variables Cualitativas Nominales y Ordinales. Asociación de la Sociedad Española de Hipertensión Liga Española para la lucha contra la Hipertensión Arterial. Abril 2004.

PITA Fernández Salvador y Pértiga Díaz Sonia. Asociación de Variables Cualitativas: test de Chi-cuadrado, Metodología de la Investigación, noviembre de 2004. [www.fisterra.com](http://www.fisterra.com)

REAL ACADEMIA ESPAÑOLA, Diccionario tomo I Vigésima primera edición, Talleres Gráficos Peñalara, S.A.,Madrid España. 1992.

RODRÍGUEZ DE VILLAQUIRAN Gilma, Estructuras Matemáticas 6, primera edición, Rei Andes Ltda. Bogota Colombia 1997 p. 312

RUNYON Richard P y Haber Audrey. Estadística para las Ciencias Sociales, 4 edición Adisson –Wesley Iberoamericana, S.A. Wilmington Delaware, E.U.A., Capitulo VIII 1992, p.122-167.

SIEGEL Sydney. Estadística no paramétrica “Aplicada a las Ciencias de la Conducta” 3ra edición, 1990 Talleres de impresora Roma S.S., Mexico Trillas. (reimpresión 1991). p. 226-273.

## 15. ANEXOS

*BATANERO Carmen, Godino Juan Díaz, Green D.R. Holmes, P Vallecillos, Errores y Dificultades en la Comprensión de los Conceptos Estadísticos Elementales.* Internation Journal of Mathematics Education in Science and Technology 1994.

En este artículo inicialmente los autores señalan la importancia de la estadística en la formación general del ciudadano, así mismo presentan un registro de los materiales (libros, revistas, proyectos, conferencias) específicos diseñados para la enseñanza por diversos países, explican que para abordar con éxito los objetivos educativos, los profesores requieren una intensa preparación en la parte disciplinar, como en los aspectos didácticos del tema lo cual incluye el conocimiento de las dificultades que se generan en la comprensión de los objetos estadísticos.

Posteriormente presentan un análisis sobre la complejidad y las dificultades para la comprensión en objetos tales como: representaciones graficas y tabulación de datos, la media, medidas de dispersión, estadísticos de orden, asociación en tablas de contingencia, diseño de experimentos, muestreo y contraste de hipótesis.

Basados en Brousseau (1983) comentan que la existencia de errores y dificultades en el aprendizaje de los conceptos básicos se puede explicar por la presencia de *obstáculos cognitivos* en los estudiantes, quien describe las siguientes características para los obstáculos cognitivos:

Un obstáculo es un conocimiento, no una falta de conocimiento. El alumno utiliza este conocimiento para producir respuestas adaptadas a un cierto contexto que encuentra con frecuencia. Cuando se usa este conocimiento fuera de este contexto genera respuestas incorrectas. Una respuesta universal exigirá un punto de vista diferente.

El alumno resiste a las contradicciones que el obstáculo le produce y al establecimiento de un conocimiento mejor. Es indispensable identificarlo e incorporar su rechazo en el nuevo saber. Después de haber notado su inexactitud, continúa manifestándolo, de forma esporádica.

Además este autor identifica tres tipos de obstáculos: *obstáculos ontogénicos, obstáculos didácticos, obstáculos epistemológicos.*

Los autores apoyados en Garfield y Alegren (1988) concluyen las siguientes razones para algunas de las dificultades que surgen en la enseñanza de la estadística:

Algunos conceptos estocásticos, tales como el de probabilidad, correlación, necesitan del razonamiento proporcional, que ha demostrado ser un tópico difícil en diversas investigaciones. Existen falsas intuiciones que los alumnos llevan consigo al empezar la enseñanza.. Aunque estas intuiciones son mejor conocidas para el caso de la probabilidad aún han sido poco estudiadas para los conceptos estadísticos.

En ocasiones los estudiantes muestran una falta de interés hacia la estadística, porque se les ha enseñado en forma muy abstracta en edades tempranas.

Dos razones mas que son señaladas por los autores son el reciente desarrollo de la estadística y que la enseñanza de los conceptos se presenta aislada de sus aplicaciones.

*BATANERO Carmen, Godino Juan Díaz y Estepa Antonio. **Construcción del Significado de la Asociación Estadística Mediante Actividades de Análisis de Datos**, Departamento de Didáctica de la Matemática, Universidad de Granada, Universidad de Jaén Madrid España. 1998*

En este artículo los autores presentan los resultados de una investigación realizada con el fin de detectar algunas problemáticas que surgen en la construcción del significado del objeto asociación estadística, concepto que tiene ligado a su aprendizaje la apropiación de tres elementos (intensionales, extensionales e instrumentales) que componen el significado institucional del objeto, teniendo en cuenta que este objeto estadístico tiene unidas a él dos dificultades una de tipo epistemológico en cuanto a que el objeto desde su disciplina es complejo y otra de tipo psicológico, ya que este concepto no se desarrolla intuitivamente.

La confrontación de este planteamiento se trabaja con dos grupos de estudiantes de primeros cursos universitarios quienes están recibiendo formación en estadística, este trabajo se realiza en dos experimentos bicitápicos, uno en el nivel descriptivo y otro en el nivel inferencial, utilizando ordenadores en las segundas fases de cada nivel. En el primer grupo la primera fase consiste en detectar los preconceptos de asociación a partir de tablas de contingencia, gráficos y diferencias de medias, destacando las concepciones erróneas a la luz de los elementos extensionales e intensionales del significado, mientras que en la segunda fase después de utilizar los ordenadores como herramientas facilitadoras (elementos instrumentales) se proponen problemas para evidenciar el aprendizaje del concepto de asociación a través de la apropiación de los tres elementos del significado antes mencionados.

En el segundo grupo se realizan las dos fases de manera similar aunque ahora las situaciones tienen que ver con objetos de la inferencia, (intervalos de confianza, test de chi cuadrado diferencia de medias).

Los autores concluyen que a través de esta investigación se evidencia la complejidad del significado de la asociación y su comprensión, destacan las concepciones erróneas que persisten después de la enseñanza con ordenadores y manifiestan la necesidad de apropiar los tres elementos mencionados para garantizar la comprensión y su adecuada utilización en la resolución de problemas.

*BATANERO, Carmen. **¿Hacia donde va la Educación Estadística?**, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España. 1999.*

En este documento la autora Carmen Batanero resalta la importancia de las investigaciones cuyo interés se enmarcan en los problemas de la enseñanza y aprendizaje de la estadística, expresa que en la actualidad existe un notable aumento de publicaciones, diseños curriculares e investigaciones relacionadas con esta temática.

El propósito de este artículo es hacer un breve resumen de este desarrollo en el ámbito internacional y reflexionar sobre la situación actual y perspectivas futuras de la educación estadística en España.

En este orden de ideas la autora presenta las razones del interés por la enseñanza de la estadística señaladas por Holmes (1980), Fischbein (1975) y más recientemente por Begg (1997). Proporciona información a los profesores e investigadores acerca de: instituciones interesadas en este tema como el IASE, sus funciones y responsabilidades, investigación sobre el razonamiento estocástico, interés por la enseñanza de la estadística dentro de la educación Matemática, como enseñar estadística y tendencias futuras en la enseñanza de la estadística, con el fin de interesarlos por la enseñanza de la estadística e impulsar el desarrollo de la investigación en este campo del conocimiento.

BATANERO Carmen. **Didáctica de la Estadística**, Grupo de Educación Estadística, Departamento de Didáctica de la Matemática, Universidad de Granada, Madrid España. 2001. p. 29-33; 00 -106.

En este documento la autora menciona que la asociación estadística es un tema que aparece con frecuencia en el análisis de datos tanto exploratorio como confirmatorio, al cual están ligadas dificultades de tipo filosófico.

Menciona que dos variables están asociadas en la extensión en que una de ellas es un buen predictor de la otra. La medida de la intensidad de la relación entre dos variables, mediante un coeficiente, constituye el problema de correlación para el caso numérico y asociación, en el caso general.

En general se refiere a la regresión y la correlación como técnicas que se preocupan de la interrelación entre dos o mas variables continuas. En la parte de regresión explica que el objeto de estudio es la relación entre una variable dependiente y otro conjunto de una o varias variables independientes, describe el desarrollo matemático inmerso en el trabajo con este concepto y menciona las razones para efectuar un análisis de regresión así como también muestra los problemas estadísticos implicados. Para la correlación puntualiza en aspectos importantes y generales tales como: papel de la asociación y la regresión en los métodos estadísticos, asociación y casualidad.

Define la causalidad desde el punto de vista estadístico, y según esta presenta el modelo determinista lineal y el modelo aleatorio lineal, en contraste con el modelo de aislamiento.

BATANERO Carmen. **Presente y Futuro de la Educación Estadística**, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España 2001A.

Carmen Batanero señala que la enseñanza de la estadística se ha ido incorporado en los distintos niveles escolares y que las instituciones interesadas por mejorar la enseñanza no sólo son los profesores, también estadísticos teóricos y prácticos aportan en este aspecto.

Resalta la labor del *International Statistical Institute* (ISI) y el *International Association for Statistical Education* (IASE) que son institutos interesados en investigar y proporcionar recursos para la educación estadística, además de apoyar la vinculación de la estadística en los currículos escolares, con el propósito de proporcionar al ciudadano lo que se conoce como *cultura estadística*, que se refiere a dos componentes interrelacionados: a) capacidad para interpretar y evaluar críticamente la información estadística, los argumentos apoyados en datos o los fenómenos estocásticos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos, y b) capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante”

Para analizar las tendencias sobre enseñanza y aprendizaje de la estadística, Batanero indica que no existen teorías específicas, por lo cual se hace necesario recurrir a las del área más cercana, la didáctica de las matemáticas, en esta parte cita diferentes teorías generales como el Constructivismo y Las Herramientas Semióticas, refiere también cuatro tipos de situaciones didácticas que son fundamentales para el aprendizaje según Brousseau (1986), (situación de acción, situación de formulación/comunicación, situación de validación, situación de institucionalización). Así mismo plantea que al preguntarse por la comprensión de los conceptos estadísticos lograda por los alumnos es fundamental analizar cuál es el significado de lo que se trata de enseñar y cuales son sus componentes. Godino (1996;1999) identifica diferentes tipos de elementos en el significado de los objetos matemáticos (el campo de problemas en donde surge el objeto, los algoritmos y estrategias empleados en la solución de problemas, representaciones, elementos conceptuales y las demostraciones).

Concluye que la educación estadística es un tema importante y prioritario que puede abordarse desde múltiples perspectivas: desarrollo de instrumentos de evaluación, identificación de concepciones iniciales y errores de los estudiantes, estudios culturales cruzados, formación de docentes, desarrollo y evaluación de recursos didácticos, evaluación de la enseñanza. Es imprescindible un mayor esfuerzo de investigación en educación estadística si se desea conseguir una mejor preparación estadística de los estudiantes.

BATANERO, Carmen. **Los Retos de la Cultura Estadística**, Departamento de Didáctica de la Matemática, Universidad de Granada Madrid España 2002.

Inicialmente la autora presenta las razones que han expuesto diferentes autores por las cuales se hace necesaria la incorporación de la estadística en el currículo desde la escuela primaria, tales como: capacitar al ciudadano para leer e interpretar tablas y gráficos, utilizar la estadística como herramienta básica y aplicable en otras disciplinas, fomentar un razonamiento crítico estadístico basado en la valoración de la evidencia objetiva y ayudar a entender los restantes temas del currículo. El propósito de esa vinculación es proporcionar al ciudadano una "cultura estadística", la cual según Gal (2002) tiene unida a ella los siguientes componentes: Conocimientos y destrezas, Razonamiento estadístico, Intuiciones y Actitudes, que son elementos que permanecen invariantes en el proceso de formación estadística.

Batanero plantea el interrogante ¿Cómo desarrollar la cultura estadística? y aborda la respuesta desde tres instancias: Los profesores, las oficinas de estadística y las sociedades estadísticas. Argumenta que, en cuanto a investigación en educación estadística en los últimos años se han presentado importantes avances desde otras ciencias del conocimiento como la psicología, la estadística y la didáctica de la matemática, pero sin duda el principal impulsor en investigación en educación estadística ha sido la Asociación Internacional para la Enseñanza de la Estadística (IASE).

Finalmente concluye que a pesar del aumento del uso de ideas estadísticas en diferentes disciplinas, los conocimientos de esta área se usan incorrectamente debido a que no se comprenden conceptos aparentemente elementales y no existe una valoración del trabajo del estadístico dentro de los equipos de investigación, esta problemática didáctica se debe a que la incorporación de la estadística desde la escuela todavía no es un hecho.

BEHAR G Roberto y Yepes A. Mario. **Estadística Un Enfoque Descriptivo**, Universidad de Valle, Facultad de Ingeniería, Departamento de Producción e Investigación de Operaciones. Talleres Gráficos de Impresora, Fenua S. A. Cali Colombia, 1996. **Capítulo III**, p. 87-151.

En este capítulo los autores inician señalando la importancia de hacer el tratamiento conjunto de dos características o variables observadas en los elementos de una muestra, para evidenciar dicha importancia plantean diversos ejemplos en los cuales, puede interesar la distribución conjunta de frecuencias, considerando conjuntamente los diferentes valores (categorías) de las variables, explican el tratamiento para estudiar las distribuciones conjuntas y distribuciones marginales según los siguientes casos: Ambas variables son de naturaleza discreta, la cual puede entenderse también como variables de tipo nominal ( ) ambas variables son de naturaleza continua una variable discreta otra continua

Consideran también el proceso para estudiar el comportamiento estadístico de una variable para los elementos que tienen un determinado valor en la otra variable considerada, lo que se conoce como: distribuciones condicionales de frecuencia.

Posteriormente muestran cuando dos variables son estadísticamente independientes o dependientes, cuando son dependientes existen grados o niveles de dependencia que deben ser mediados de tal manera que permita poner en evidencia la intensidad de la dependencia

estadística (asociación), para lo cual presentan algunos indicadores como el cuadrado de contingencia  $\chi^2$  chi – cuadrado, el cuadrado medio de contingencia y el coeficiente de contingencia.

Finalmente presenta dos nuevos elementos de asociación estadística: covarianza y coeficiente de correlación lineal, para detectar o conocer sobre la fuerza de asociación estadística de dos variables en la dirección de una línea recta.

BEHAR G Roberto y Yepes A. Mario. **Estadística Un Enfoque Descriptivo**, Universidad de Valle, Facultad de Ingeniería, Departamento de Producción e Investigación de Operaciones. Talleres Gráficos de Impresora, Fenusa S. A. Cali Colombia, 1996. **Capítulo IV**, 157-209.

Roberto Behar y Mario Yepes en el capítulo “El Modelo de Regresión” presentan el modelo de regresión lineal desde su origen con Sir Francis Galton, quien en sus estudios biológicos sobre herencia, encontró que las características de una generación tendían a moverse en dirección del promedio de la característica de la población, a este fenómeno lo denoto regresión filial.

Los autores muestran la importancia de la asociación de variables y la regresión lineal simple, y relacionan algunas aplicaciones.

Presentan el proceso para generar la línea de regresión desde el diagrama de dispersión y la línea mínimo cuadrática, donde se incorpora el coeficiente de correlación y la búsqueda de los parámetros para ajustar el mejor modelo, con el fin de realizar predicciones confiables. Llegando a obtener las ecuaciones normales, hace alusión también a como debe interpretarse el modelo de regresión y presentan algunas observaciones pertinentes para la interpretación de los parámetros. Posteriormente hacen la generalización para el proceso de estimación de mínimos cuadrados, explica los beneficios y debilidades, construye un indicador de bondad de ajuste de un modelo. Presenta además algunas familias no lineales, y la transformación adecuada para linealizarlos.

CANAVOS George C. **Probabilidad y Estadística Aplicación y Métodos**, Virginia Commonwealth University, McGraw-Hill / Interamericana de México S.A. De C. V. 1988. **Capítulo XIII** p. 443 – 571.

En el capítulo “Análisis de Regresión: El Modelo Lineal Simple”, el autor examina las asociaciones cuantitativas entre un número de variables lo que se conoce como: *análisis de regresión*, para el cual presenta dos significados: el primero surge de la distribución conjunta de probabilidad, el segundo nace de la necesidad de ajustar un modelo a un conjunto de datos. Para este último, propone y explica el método de los mínimos cuadrados para el modelo lineal simple en el que sólo se tiene una variable de predicción y se supone una ecuación de regresión lineal de la forma:

$Y_i = \beta_0 + \beta_1 x_i + e_i$  para  $i = 1, \dots, n$ , donde  $Y_i$ : es la  $i$ -ésima observación de la variable respuesta, la cual corresponde al  $i$ -ésimo valor  $x_i$  de la variable de predicción,  $e_i$  es el error aleatorio asociado con  $Y_i$ ; y  $\beta_0$  y  $\beta_1$  son los parámetros desconocidos. El objetivo de la técnica de los mínimos cuadrados es obtener estimaciones para estos parámetros. El autor formula algunos supuestos para los estimadores por mínimos cuadrados que permiten obtener la línea de regresión.

$\hat{Y}_i = B_0 + B_1 x_i$ , donde  $B_0$  y  $B_1$  son los estimadores para  $\beta_0$  y  $\beta_1$  respectivamente.

Presenta las propiedades generales de los estimadores considerando criterios que permitan la construcción de intervalos de confianza y la realización de pruebas de hipótesis con respecto a los parámetros del modelo de regresión simple, para lo cual hace uso de la inferencia estadística y del análisis de varianza. Además de la técnica por mínimos cuadrados Canavos emplea el principio de

máxima verosimilitud y el álgebra de matrices para estimar los parámetros del modelo lineal simple, obteniendo los mismo estimadores que con el método de los mínimos cuadrados.

CASTILLO M Alberto y Ojeda R. Mario Miguel, **Principios de Estadística no Paramétrica**, 1ra edición, Universidad Veracruzana, Xalapa, Veracruz., México, octubre de 1994. p.143-149.

El autor hace referencia que cuando se tienen observaciones aparejadas de un fenómeno, es natural preguntarse si existe alguna relación entre las variables, si existe dicha relación esta se detectará a través del coeficiente de correlación, el signo indica si la relación es directa o inversa, y la intensidad está dada por la magnitud del coeficiente. Hace notar que el coeficiente de correlación no indica relación causal entre las variables, cuando estas no tienen una relación lineal.

Presenta algunas medidas de asociación tales como; el  $r$  de Pearson, el coeficiente de correlación  $\rho$  de Spearman, Phi  $\Phi$ , V de Cramer, coeficiente de contingencia, lambda asimétrica, lambda simétrica, describe sus características principales, sus bondades y hace referencia a las condiciones para utilizar una u otra de estas medidas.

CAZU Pablo. **La Investigación Bibliográfica**, Buenos Aires Argentina, marzo 2000.

Pablo Cazau, describe la investigación bibliográfica como la etapa de la investigación científica donde se explora qué se ha escrito en la comunidad científica sobre un determinado tema o problema, presenta algunas indicaciones acerca de qué consultar y cómo hacerlo, describe las cinco etapas de la investigación científica: *a. Definir el tema, b. Revisión Bibliográfica alrededor del tema, c. Trazar un Proyecto, d. Ejecutar el Proyecto, e. Exponer los resultados*. Especifica algunas ventajas de realizar investigación bibliográfica. Dentro de este artículo presenta y describe los tres niveles de bibliografía a consultar y como consultarla de acuerdo al tipo de destinatario para el que fue diseñado: a. Bibliografía para público en general, b. Bibliografía para aprendices o alumnos, c. Bibliografía para profesionales e investigadores, recomienda como realizar citas bibliograficas lícitas, menciona y describe la propuesta de clasificación de las fuentes de información de Dante y presenta algunas recomendaciones de otros autores al respecto. Recomienda registrar la información recolectada en fichas y destaca actividades propias de este tipo de investigación.

ESTEPA Antonio y Sánchez Francisco. **Correlación y Regresión en los Primeros Cursos Universitarios**. Departamento de Didáctica de las Ciencias, Departamento de Matemáticas, Escuela Politécnica Superior. Universidad de Jaén. Madrid. España. 2000

En la introducción de este artículo los autores señalan la importancia de la estadística en la sociedad actual, lo cual ha generado una incorporación de esta área en los currículos de educación, lo que conlleva a la necesidad de realizar investigación en didáctica; determinan que uno de los temas de importancia en estadística es el concepto de asociación, tanto en sí mismo como por ser prerrequisito previo para continuar avanzando en el conocimiento estadístico.

Realizan una revisión bibliográfica, donde abordan los resultados de investigaciones psicológicas y didácticas que se han realizado para la enseñanza del tema, una problemática que emerge de las investigaciones psicológicas es la escasa capacidad que tienen los adultos para estimar la correlación, mientras que en las investigaciones didácticas se detectan cuatro concepciones erróneas sobre asociación estadística (concepción determinista, concepción unidireccional, concepción local y concepción causal).

Apropian el marco teórico sobre el Significado Personal e Institucional de los Objetos Matemáticos propuesto por Godino (1994) para realizar su experimento de enseñanza de la asociación estadística, el cual se realiza sobre un grupo de 193 estudiantes universitarios de cursos introductorios en estadística con el propósito de caracterizar el significado de algunos elementos de

la correlación y regresión. El cuestionario propuesto contiene items en los cuales a partir de diferentes representaciones de datos y dos problemas propuestos se pide al alumno: calcular el coeficiente de correlación, decir que tipo de relación (directa, inversa o independencia) existe entre las variables, determinar las rectas de regresión, usar la recta de regresión para predecir.

Finalmente presentan los resultados obtenidos alrededor de las prácticas realizadas, además argumentan que estos, necesitan mucha investigación para quedar clarificados en su totalidad, pero que si se tienen en cuenta para la planificación de la enseñanza esta mejorará.

FERRIS J. Ritchey, **El Potencial de la Imaginación Estadística**, Mc Graw Hill 2002 p. 365 –371

En este documento el autor expone que una vez detectada una relación entre dos variables se puede considerar sobre esta cuatro aspectos: existencia, dirección, fuerza y naturaleza. La existencia de una relación entre dos variables para todos los sujetos de una población se determina con base en el análisis estadístico de una muestra, si no existe relación entre dos variables, entonces la relación no tiene ni dirección, ni fuerza, ni naturaleza.

La dirección de una relación puede ser positiva o negativa, una relación positiva es aquella en la cual un incremento en una variable implica un incremento en la otra, una relación negativa es en la cual un incremento en una variable se relaciona con un decremento en la otra.

La fuerza de una relación entre dos variables establece el grado en que se reducen los errores al predecir las puntuaciones de una variable dependiente.

La naturaleza tiene que ver en como ayuda el conocimiento de una relación entre dos variables para entender y predecir los resultados de la variable dependiente.

Posteriormente señala cuales aspectos de una relación son útiles para cada prueba bivariada.

GODINO Juan Diaz y Batanero Carmen. **Significado Institucional y Personal de los Objetos Matemáticos**, Recherches en Didactique des Mathématiques, Vol 14 Revista No. 3 p. 325 – 355. 1994

En este trabajo los autores proponen una teoría pragmática acerca del significado de los objetos matemáticos. Para desarrollarla, abordan el análisis de la noción del significado desde un punto de vista didáctico, es decir desde los procesos de enseñanza y aprendizaje y lo relacionan con los usos que se le dan a los objetos matemáticos en el contexto en donde son manipulados, esto último es lo que permite que la teoría adquiera su posición pragmática.

De acuerdo a lo anterior los autores empiezan definiendo la noción de práctica como toda actuación del individuo o la institución (grupo de personas) encaminada a resolver un problema matemático, y la diferencian de práctica significativa atendiendo al hecho de que la resolución de un problema matemático no necesariamente responde a un proceso lineal ya que es frecuente que se presenten intentos fallidos. Así se dice que una práctica es significativa si desempeña una función para la consecución del objetivo en los procesos de resolución de un problema, o bien para comunicar a otro la solución, validar la solución y generalizarla a otros contextos y problemas.

Las nociones de Práctica y Práctica Significativa son la base para edificar toda esta teoría y llegar así a construir el significado de los objetos matemáticos, tanto desde lo personal como de lo institucional y así poder realizar comparaciones entre estos dos significados, lo cual servirá para formular criterios de evaluación del aprendizaje.

MOLINERO Luis M. **Asociación de Variables Cualitativas Nominales y Ordinales**. Asociación de la Sociedad Española de Hipertensión Liga Española para la lucha contra la Hipertensión Arterial. Abril 2004.

En este artículo el autor menciona el hecho de que por la rutina el tratamiento de los datos estadísticos referidos al concepto de tablas de contingencia con frecuencia se resume a la tarea simple de introducir datos en un programa informático limitándose a transcribir mecánicamente los resultados lo cual conlleva a no realizar un mayor análisis y restringir nuestra mirada solo a los resultados con los que estamos familiarizados, y olvidándonos del resto de información que quizás no entendemos.

Para el desarrollo de este documento el autor se basa en la pregunta ¿Cuál de estos test: Pearson chi<sup>2</sup>, Likelihood-ratio chi<sup>2</sup>, Cramér's V, Gamma, Kendall's tau-b, me indican que a peor estado Hoehn y Yahr haya más pacientes con fatiga?, realizada con respecto a una tabla producto de una salida de un programa estadístico. En este sentido explica el funcionamiento del chi<sup>2</sup> de Pearson para la prueba medica describe el algoritmo general para su calculo por ser este el mas utilizado, con relación al resto de coeficientes considerados describe lo que significan cada uno de ellos y por último se refiere a la importancia de las pruebas de significancia.

PITA Fernandez Salvador y Pértiga Díaz Sonia. **Asociación de Variables Cualitativas: test de Chi-cuadrado**, Metodología de la Investigación, noviembre de 2004. [www.fisterra.com](http://www.fisterra.com)

Se presenta el análisis de los indicadores de asociación arrojados por una programa estadístico de una tabla de contingencia que relaciona los datos observados de dos variables en un estudio medico.

En este artículo se presentan las tablas de contingencia como herramienta útil de la estadística para el procesamiento de datos, advierte que deben usarse con cuidado y meticulosidad, y no subvalorar información para el análisis de datos.

El autor destaca los indicadores de asociación de Chi<sup>2</sup> de Pearson y el chi<sup>2</sup> del cociente de máxima verosimilitud, como asintóticamente equivalentes

Puntualiza que se entiende por asociación y por significación no desconocer principalmente de variables cualitativas, nominales y ordinales, con el fin de detectar una relación entre ellas.

La técnica en primera instancia consiste en formular la prueba de hipótesis, como hipótesis nula se considera que las variables no estén asociadas y como hipótesis alterna que lo estén, bajo el supuesto de que la hipótesis nula se cumple se hace uso del estadístico Chi – cuadrado en donde intervienen las frecuencias observadas y las frecuencias esperadas para su calculo, teniendo en cuenta los grados de libertad como parámetro y los valores estandarizados en la distribución chi-cuadrado se toma la decisión de rechazar la hipótesis nula según el valor de este estadístico sea grande.

En este artículo los autores mencionan que en la investigación bioestadística se encuentra con frecuencia con datos o variables de tipo cuantitativo, mediante las cuales un grupo de individuos se clasifican en dos o más categorías excluyentes.

Presentan un ejemplo de tablas de contingencia 2x2 y hacen el desarrollo de la prueba X<sup>2</sup> (en el contraste de independencia de variables aleatorias cualitativas) para este tipo de tablas, además de que muestran y describen el algoritmo general para tablas de contingencia de tamaño rxk (r filas y k columnas).

Por ultimo aparece consignada en una tabla los valores de la distribución de probabilidad de ji-cuadrado con uno hasta cien grados de libertad

RUNYON Richard P y Haber Audrey. **Estadística para las Ciencias Sociales**, 4 edición Adisson – Wesley Iberoamericana, S.A. Wilmington Delaware, E.U.A., Capitulo VIII 1992, p.122-167.

Los autores exponen que cuando se necesita determinar las relaciones entre dos o mas variables, es necesario incursionar en el área de la correlación, donde se puede encontrar una medida cuantitativa del grado en que dos variables están relacionadas o tienden a variar conjuntamente, la intensidad del coeficiente viene dado en términos de la magnitud. Existen varios coeficientes de correlación y determinar cual es el adecuado para el problema depende de factores como:

Tipo de escala en la cual están medidas las variables  
La naturaleza de la distribución (continua o discreta)  
La característica de la distribución (lineal o no lineal)

En este capitulo los autores abordan el estudio de dos coeficientes: la  $r$  de Pearson, que se usa con datos en escalas de intervalo o de razón; y el  $r$  de Spearman que se usa con variables ordinales, presentan una tabla que ilustra los diversos tipos de coeficientes y las escalas con que son utilizados, muestran algunas características comunes en la obtención de un coeficiente, a través de la obtención de dos conjuntos de mediciones en los mismos individuos (o sucesos) o en parejas de individuos que tengan alguna forma de relación

Presentan las propiedades generales de los coeficientes, intervalo de variación, y tipos de relación de acuerdo a la magnitud del coeficiente y cuando se detecta independencia lineal o ausencia de relación

Presentan también varios diagramas de dispersión en los cuales ilustran los diferentes grados de relación entre las variables

Presentan la relación entre la  $r$  de Pearson y los valores  $z$ , que representan una medida de posición relativa en una variable dada, de este modo la  $r$  de Pearson es una medida del grado en que los mismos sucesos ocupan la misma posición relativa respecto de los valores.

La  $r$  de Pearson es apropiada exclusivamente para variables relacionadas de una manera lineal. Con datos ordenados, el coeficiente de correlación por rangos de Spearman

SIEGEL Sydney. **Estadística no paramétrica “Aplicada a las Ciencias de la Conducta”** 3ra edición, 1990 Talleres de impresora Roma S.S., Mexico Trillas. (reimpresión 1991). Capitulo IX, p. 226-273.

El capitulo Las Medidas de Asociación y sus Pruebas de Significación esta dedicado a las medidas de correlación no paramétricas, los autores señalan el interés de indicar la medida del grado de asociación entre dos conjuntos de valores de un grupo dado de individuos. Pero es quizás de mayor interés afirmar si alguna asociación observada en una muestra de puntajes indica que la variable en estudio están asociadas muy probablemente en la población en que se tomo la muestra. De acuerdo a lo anterior presentan las siguientes cinco técnicas no paramétricas para medir el grado de correlación entre las variables de una muestra: coeficiente de contingencia, coeficiente de correlación de rango de Spearman, coeficiente de correlación de rango de Kendall, coeficiente de correlación parcial de rango de Kendall y coeficiente de concordancia de Kendall. Para cada una de ellas a excepción del coeficiente de correlación parcial de Kendall muestran las pruebas estadísticas que determinan la “significación” de la asociación observada, las cuales pueden emplearse con datos nominales y ordinales. Las pruebas suponen la forma de la población de la cual se extrajeron los puntajes.

## **Variables estadísticas bidimensionales (tema 4)**

Autores:

En este capítulo el autor explica que es interesante conocer el comportamiento de dos variables  $X$  y  $Y$  en una población cuando se realizan estudios estadísticos. Para algunos fenómenos es posible encontrar una fórmula que explique una variable en términos de la otra, estos fenómenos son llamados *deterministas*, pero existen fenómenos que no se pueden explicar mediante una relación funcional. Para este último caso se estudia la asociación entre las variables según la naturaleza de las mismas, existen tres tipos de estudio:

El estudio de la asociación entre variables cualitativas se hace mediante tablas de contingencia.

En el caso de una variable cuantitativa y una cualitativa se estudia la asociación analizando la diferencia de las dos medias y comparando los intervalos de confianza de las dos medias.

Cuando las variables son cuantitativas se habla de correlación entre las variables.

Con el propósito de realizar un estudio sobre algunos elementos de la asociación, el autor inicia presentando la distribución conjunta de dos variables estadísticas a partir de las tablas de contingencia, mostrando que se pueden obtener además distribuciones unidimensionales cuando se condiciona alguna variable. Indica que el mayor interés de las distribuciones condicionadas es el estudio de la dependencia aleatoria.

Más adelante para el caso de variables numéricas presenta la covarianza y el coeficiente de Pearson como coeficientes cuyo valor indica el tipo de relación entre las variables, cuando existe suficiente correlación entre dos variables numéricas se plantea el problema de determinar la ecuación matemática que permite predecir una de las variables en términos de la otra, esta función será la línea de regresión de  $Y$  en función de  $X$ .

Por otra parte el autor expone que existen herramientas en statgraphics para realizar tablas bidimensionales para dos variables cualitativas y obtener las distribuciones condicionadas por filas y por columnas, además de diferentes opciones gráficas para en el estudio del razonamiento sobre la asociación estadística representar las distribuciones, con este mismo programa se puede estudiar la regresión y la correlación entre las variables.

En la parte final del documento comenta que la investigación sobre el concepto de asociación ha sido objeto de gran interés en la psicología y ha estado ligada a los estudios sobre toma de decisiones, presenta las dificultades que encuentran investigadores pioneros en el estudio del razonamiento sobre la asociación estadística, en experimentos realizados.