



Universidad Nacional Autónoma de México

FACULTAD DE PSICOLOGIA

(22/5)

ESTIMACION BAYESIANA DE CALIFICACIONES

Tesis Profesional

Que para obtener el Título de
LICENCIADA EN PSICOLOGIA

p r e s e n t a

ADELINA ROSAS MERCADO

Maestro Asesor: **DR. LUIS CASTRO**

Noviembre de 1983



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

75053.08

UNAM. 103

1983

ej. 2.

M-20217

lps. 1055

A mis padres.

3111

CONTENIDO

Resumen.

vi

I.-INTRODUCCION.

1-1	BREVE HISTORIA DE LA MEDICION EDUCATIVA.	1
1-1.1	De Antes de Cristo a la Edad Media	2
1-1.2	El Siglo XIX.	4
1-1.3	El Siglo XX.	6
1-1.31	Confiabilidad del calificador.	7
1-1.32	Desarrollo de las pruebas de aprovechamiento.	7
1-2	MEDICION Y EVALUACION EN EL PROCESO EDUCATIVO	11
1-2.1	Escala Nominal.	12
1-2.2	Escala Ordinal.	12
1-2.3	Escala de Intervalos.	13
1-2.4	Escala de Razón.	14
1-2.5	Medición y Evaluación.	15
1-2.6	Problemas de la Medición y de la Evaluación.	16
1-3	LA IMPORTANCIA DE LA CORRECTA ESPECIFICACION DE OBJETIVOS EN LA EVALUACION.	17
1-4	FORMULACION DEL PROBLEMA DE ASIGNACION DE CALIFICACIONES.	21
1-4.1	Significado de las Calificaciones.	34
1-4.2	Marco de Referencia para la Asignación de Calificaciones.	35

1-4.21	Ejecución en relación con la perfección.	35
1-4.22	Ejecución en relación con el potencial	36
1-4.23	Ejecución en relación con los iguales.	36

II.- EL METODO BAYESIANO.

2-1	INTERPRETACION DE PROBABILIDADES.	40
2-1.1	Principios de Probabilidad.	43
2-1.2	La Teoría Clásica.	44
2-1.3	La Teoría de Frecuencia Relativa.	45
2-1.4	La Teoría Personalista.	47
2-1.41	Bayes. Consideraciones preliminares.	48
2-2	CONCEPTOS BASICOS SOBRE PROBABILIDAD.	51
2-2.1	Primer Axioma.	52
2-2.2	Segundo Axioma.	53
2-2.3	Tercer Axioma.	53
2-2.4	Cuarto Axioma.	53
2-2.5	Probabilidad de un evento	54
2-3	PROBABILIDAD CONDICIONAL.	54
2-4	TEOREMA DE BAYES.	55
2-5	ALGUNOS USOS DE LA ESTADISTICA BAYESIANA	62

III.-METODO PROPUESTO.

3-1	ANTECEDENTES.	65
3-2	PROCEDIMIENTO	69

I.-Determinación de la Muestra.	73
II.-Determinación de Frecuencias de Calificaciones para la Muestra.	77
III.-Determinación de Probabilidades a Priori	77
IV.-Determinación de las Probabilidades de Intersección	78
V.-Cálculo de Valores Esperados Tradicional v Bayesiano.	82
VI.-Calificaciones Iniciales v Definitivas para la Población.	85
VII.-¿Es el Teorema de Baves un Mejor Método de Estimación de la Calificación Definitiva que el Estimador de Probabilidad Clásica Tradicional?	86
VII.1 Análisis gráfico.	86
VII.2 Prueba de significancia.	90
VIII.-Conclusiones	94

TABLAS.

Tabla 1 Porcentaje de calificaciones. Errores.	32
Tabla 2 Tamaño del error.	33
Tabla 3 Aspirantes a ingreso en una universidad.	55
Tabla 4 Elementos de la muestra según calificación inicial (C_I) v calificación definitiva (C_D)	77

Tabla 5	Probabilidades de intersección de las calificaciones de la muestra.	79
Tabla 6	Probabilidades a posteriori.	81
Tabla 7	Elementos de la población según calificación inicial (C_I) v calificación definitiva (C_D)	
Tabla 8	Frecuencias observadas v esperadas.	89
Tabla 9	Frecuencias observadas v esperadas (para dos categorías).	90
Nomenclatura.		95
Apéndice 1 Glosario.		97
Referencias bibliográficas.		109

Resumen.

Utilizando la teoría de la inferencia bayesiana como marco de referencia y la teoría de probabilidad clásica como base de comparación, se propone un procedimiento estadístico para la predicción de calificaciones o notas que pretende: dada una calificación inicial - asignada en función de ejecuciones evaluadas en relación con los criterios especificados en los objetivos de una materia-, estimar la calificación final. Ambos procedimientos fueron aplicados a una misma muestra de sujetos (60) y las probabilidades inferidas se compararon con los datos de la población (correspondiente a 332 registros de calificaciones de un mismo profesor, en una escuela de ingeniería del IPN). Los resultados de un análisis gráfico y de una prueba χ^2 , mostraron que el estimador bayesiano fue más preciso a un nivel de significancia menor de 0.05.

ESTIMACION BAYESIANA DE CALIFICACIONES.

ADELINA ROSAS MERCADO

El presente trabajo propone un procedimiento estadístico, basado en el Teorema de Bayes, para la predicción de calificaciones o notas; este método pretende: dada una calificación inicial, predecir la calificación final.

Tanto el método bayesiano propuesto, como otro de probabilidad clásica, serán aplicados a una misma muestra de sujetos. Se predecirán las calificaciones finales esperadas con cada uno de los procedimientos. Posteriormente, por un método gráfico, se analizará cada predicción. Finalmente, se aplicará un procedimiento apropiado para prueba de hipótesis en casos semejantes y se discutirán los resultados.

Se iniciará el estudio con una brevísima historia de la medición y evaluación educativas, con el fin de aportar claridad a los conceptos de medición y evaluación que en seguida se consignan. El tratamiento estadístico que aquí se propone, atiende exclusivamente al proceso de evaluación. En seguida, se toca someramente el problema de la correcta especificación de objetivos y su importancia. Estos preliminares tratan de enfocar y clarificar la estructura de los factores que conforman el problema de la asignación de calificaciones. A continuación se mencionan los principales estudios realizados sobre metodología bayesiana en evaluación educativa, con lo que los antecedentes ne

cesarios para la formulación del problema que en seguida aparece habrán sido enunciados.

También se exponen, los principales lineamientos de los procedimientos bayesianos y el método que atendiendo a todos los factores mencionados, se propone como opción de solución al problema.

Finalmente, se realiza un análisis sobre lo propuesto. Con objeto de no desviar la atención del tema principal, se anexa en la parte complementaria del informe un glosario en el que se explican los términos más importantes utilizados.

1-1 BREVE HISTORIA DE LA MEDICION EDUCATIVA.

1-1.1 De antes de Cristo a la Edad Media.

Los orígenes de la medición y evaluación educativos se encuentran perdidos en la historia. Sin embargo, se tienen noticias de que en China, varios siglos antes de Cristo, ya existía un sistema bastante elaborado de pruebas de servicio social, así como sistemas de equivalencias y certificación de estudios. (1)

En Grecia, la evaluación por medio de exámenes era un sistema comunmente usado en el proceso educativo. La mayéutica socrática - método de evaluación y de enseñanza con realimentación inmediata- mejorada por Platón con su dialéctica, utiliza un tipo de examen oral para la investigación de ciertos conocimientos, habilidades y actitudes sobre tópicos particulares; Pla

(1) DU BOIS, P.H.A. 1970. "History of psychological testing, Boston: Allyn and Bacon Inc.

tón establece también en su Academia una evaluación diagnóstica, basada en una premedición educativa que realizaba a todos los aspirantes a ser admitidos en su escuela. (A la entrada de la Academia, se encontraba un letrado que decía: "No entre aquí quien no sepa geometría"). (2)

En la Biblia, se encuentran algunos registros de ciertos sistemas de evaluación en los que la persona que no pasaba la prueba era condenada a muerte (3) .

En la Edad Media, existían objetivos educativos precisamente definidos; cuando el aprendiz entraba a los gremios o corporaciones, si deseaba obtener el grado de "maestro", necesitaba desarrollar una obra maestra exactamente igual a una realizada - por el maestro del taller. Se tiene, además de objetivos precisamente definidos, el criterio de ejecución aceptable del 100 %; he aquí un caso de evaluación referida a un criterio en el dominio psicomotor. (Stanley y Hopkins consideran que es precisamente en el dominio psicomotor donde la evaluación referida a un - criterio cumple ampliamente sus funciones.) (4) .

(2) GUTRIE, W.K.C. "Los filósofos griegos". México: Breviarios F.C.E.. 1955.

(3) JUECES, 12: 5-6, Versión del Rey Juan.

(4) STANLEY, J.C and K.D. HOPKINS. 1976. "Educational and Psychological Measurement and Evaluation, Englewood Cliffs, N.J.: Prentice-Hall, Inc.

La utilización que se ha venido haciendo de la medición y evaluación educativas, tal vez ha dependido siempre de la cosmovisión de los individuos, de los valores contextuales de la sociedad y de los propósitos particulares de las personas que realizan las mediciones y las evaluaciones. Los ambientes sociales cambian a lo largo de la historia, así como los matices de las formas y criterios de evaluación. Son los intereses de los grupos que detentan el poder económico político, los que quizá determinen en cierta medida qué se va a estudiar, quién va a estudiar y en consecuencia qué se va a evaluar y cómo.

1-2.2 El siglo XIX.

Stanley y Hopkins (1976), Thorndike R y E. Hagen (1970) y Du Bois (1970), recogen los siguientes hechos significativos en la historia de la medición y evaluación educativas: la evaluación educativa anterior al siglo XIX estaba basada en pruebas orales; a pesar de que a mitad de siglo se disponía de papel y lápices accesibles, los exámenes continuaron siendo básicamente orales.

Horace Mann (1845) en Nueva Inglaterra, expuso la importancia de las pruebas en la educación y señaló que las que estaban utilizándose sufrían de ciertos defectos y limitaciones: el maestro evaluaba subjetivamente, no había uniformidad ni en las preguntas formuladas, ni en la evaluación de las respuestas; la comparación que se realizaba entre diferentes alumnos no era adecuada ni práctica, pues sólo se podía examinar a un alumno cada vez y esto llevaba mucho tiempo; además, el nerviosismo del alumno y la actitud del maestro al preguntar influyen en el de-

sempañío del examinado.

El siglo XIX presenció grandes avances en la física, química, fisiología y neurología, grandes descubrimientos empíricos y la aparición de instrumentos de medición -cada vez más precisos- y de métodos de investigación. La psicología, que en esos tiempos formaba parte de la filosofía, se hallaba sumergida en un mar de especulación; con base en una observación superficial de los fenómenos se habían desarrollado infinidad de teorías psicológicas. El inicio de la medición en psicología recibió influencias de ciertos descubrimientos en la biología y en la física: Charles Darwin publica: "El origen de las especies" en 1859, enunciando entre otros aspectos, que la vida es la lucha por la existencia y únicamente los individuos mejor dotados tienen mayores probabilidades de sobrevivir y reproducirse. Francis Galton, seguidor de Darwin, estudió cuidadosamente las diferencias entre los hombres y en su obra: "Genio hereditario", aparecida en 1869, aplicó el método estadístico a la genética. Sus estudios sobre las diferencias individuales devienen en el desarrollo de los tests de medición y del sistema de calificación "por curva" de los rendimientos de los alumnos. Asimismo Pearson, alumno de Galton, desarrolló procedimientos estadísticos de importancia fundamental en investigación psicológica y educativa.

En 1864, George Fisher propuso el uso de medidas objetivas y estandarizadas para logros académicos.

Emerson E. White, en 1886 decía en Estados Unidos, que era un hecho establecido el que la instrucción y el estudio que

se desarrollaban en la escuela no podían ser nunca mejores que la calidad de las pruebas con que se medían los desempeños de los alumnos, hablaba también de las ventajas del examen escrito en contraposición con el oral.

La estandarización de las pruebas fue promovida por Joseph M. Rice en 1897. Se evidencia así, que en la Unión Americana se realizaron tareas importantes en lo concerniente a medición y evaluación educativas; no obstante, estas ideas no se difundieron sino tiempo después.

1-1.3 El siglo XX.

1-1.31 Confiabilidad del calificador.

Max Meyer (1908) encontró diferencias en la confiabilidad del calificador en diferentes estudios realizados. Daniel Starch y Edward Elliot (1913) analizaron las calificaciones dadas a la misma prueba de geometría por 116 maestros de matemáticas de secundaria y encontraron que el rango de las calificaciones iba del 23 al 92 por ciento. Con esto se hace patente que con una inadecuada o nula especificación de objetivos, defectos técnicos en la construcción de pruebas y desconocimiento de los factores que decrecientan la confiabilidad de las calificaciones, ni siquiera los maestros de matemáticas se ponen de acuerdo en lo referente a la asignación de puntajes a las pruebas.

Starch en 1913 encontró que aún los mismos maestros daban diferentes calificaciones a las mismas pruebas en ocasiones diferentes.

Falls en 1920 realizó un estudio semejante con 100 ma-

estros de inglés que calificaron una composición. Los valores asignados variaron de 60 a 98 con un promedio de 87 %.

Ashbaugh en 1924 realizó la misma experiencia tres veces, con intervalos de cuatro semanas y encontró variaciones de 8.1 puntos entre el primer y el segundo ensayo y de 7.3 en promedio entre el tercero y el cuarto.

Hulten en 1925, en un estudio similar, encontró que - las calificaciones asignadas a una prueba de composición por 29 maestros de inglés, variaban ampliamente después de un intervalo de 2 meses.

1-1.32 Desarrollo de las pruebas de aprovechamiento.

El movimiento moderno de las pruebas de aprovechamiento fue iniciado por Thorndike en 1904 al publicar su libro de texto sobre medición educativa. En 1910 publicó la escala sobre escritura a mano que estaba formada por muestras de escritura formal de niños de quinto a octavo grado. Las muestras se arreglaron en unidades iguales con 15 categorías ordinales de calidad.

C.W. Stone en 1908 realizó pruebas estandarizadas de aritmética en primaria. De 1904 a 1920 se realizó un desarrollo fundamental en calificación objetiva: los instrumentos estandarizados de aprovechamiento en diferentes áreas. En 1915 la Asociación Nacional de Investigación Educativa surgió a la luz y ya para 1918 había una gran cantidad de bibliografía sobre tests con la nueva orientación.

Los investigadores, así pues, se empezaron a preocupar por mejorar los tipos de reactivos para elaborar pruebas que midieran habilidades "mentales altas"; se desarrollaron en consecuencia nuevos tipos de reactivos para medir habilidades como: análisis crítico, entendimiento y comprensión (cualesquiera cosas que sean). Surge entonces, un énfasis sobre la confiabilidad y validez de las pruebas estandarizadas y nuevos métodos estadísticos y técnicas de investigación fueron desarrollados para dar respuesta a esas y nuevas preguntas.

En la época de los veinte se publicaron muchos libros de texto con un suplemento sobre el trabajo de Thorndike: "Introducción a la teoría de las mediciones mental y social".

Entre 1930 y 1940 el número de tests desarrollados, estandarizados y publicados fue enorme. Se crearon organizaciones al respecto, y para 1940, se disponía de 2600 pruebas de rendimiento sobre lectura, lenguaje, matemáticas, ciencia, salud, comercio, aeronáutica e ingeniería.

La publicación en 1936 de los métodos psicométricos de Guilford, tuvo un impacto muy grande, llamando la atención sobre los aspectos teóricos y técnicos de la medición. En esta época aparecieron numerosos tests, además de la máquina IBM 805, desarrollada para asignar puntajes a las hojas de respuesta. La aparición de esta máquina facilitó el análisis de reactivos.

En 1942 Smith y Tyler en un reporte de la Asociación de Educación Progresiva, declararon que la medición educativa había sobreenfatizado habilidades y áreas de conocimiento y excluido obje

tivos educativos de suma importancia; puntualizaron que los objetivos educativos se habían realizado para medir cambios en la conducta de los alumnos y que una evaluación valiosa debería desarrollar los medios adecuados para medir amplias áreas de aprendizaje. En consecuencia, estos investigadores desarrollaron procedimientos para medir habilidades como pensamiento crítico, sensibilidad social y ajuste personal. El movimiento de Lindquist y Tyler tuvo una influencia profunda en la medición educativa ya que condujo a una mejora en la medición de procesos "mentales elevados", tales como habilidades intelectuales, intereses y actitudes.

Es un hecho significativo observar que la Primera y Segunda Guerras Mundiales, estimularon el desarrollo de prácticas de medición, pues unificaron muchos esfuerzos de investigación bajo los auspicios del gobierno.

A partir de la Segunda Guerra Mundial, ha habido un esfuerzo continuo por incrementar la precisión y utilidad de los instrumentos de medición en educación y en psicología. La precisión en la aplicación del método científico se incrementó aún más con la aparición al final de los treinta y principio de los cuarenta del *Psychometrika* y del *Educational and Psychological Measurement*, publicaciones periódicas, dedicada al desarrollo de la psicología como ciencia cuantitativa y racional, la primera; y al desarrollo y aplicación de medidas de diferencias individuales la segunda.

En 1955 fueron publicadas: "Recomendaciones Técnicas para las Pruebas de Aprovechamiento" producidas por el NCME (Consejo Nacional de Médiación en Educación) y el AERA (Asociación A-

americana de Investigación Educativa). Este documento tuvo importancia sobre los estándares utilizados en tests publicados.

En 1956 es publicada la taxonomía de objetivos de Bloom y colaboradores, teniendo ésta, una importancia fundamental en la calidad de los objetivos y de los reactivos de las pruebas objetivas. En 1959, el libro de Robert F. Mager sobre redacción de objetivos conductuales, publicado en edición pirata por Cuba, establece las bases para el desarrollo de metas claras y de significado casi unívoco, así como de criterios de ejecución y establecimiento de condiciones para la medición de rendimientos de alumnos.

Resumiendo, se tiene que antes de 1850 había un uso muy extendido de exámenes orales, teniendo que la predicción involucrada en la evaluación no era cuantificable ni comprobable.

Entre 1850 y 1900 se empieza a gastar, por Horace Mann, George Fisher, Francis Galton, Joseph M. Rice y otros, el cambio de las pruebas orales por las escritas, incrementándose los niveles de confiabilidad de las calificaciones.

Entre 1900 y 1935, el libro de texto de Thorndike, las primeras pruebas estandarizadas: Otone (aritmética), Thorndike (escala de escritura a mano) produjeron un gran avance en la medición educativa.

Entre 1925 y 1950 se desarrollaron una gran cantidad de pruebas de aprovechamiento estandarizadas y entre 1950 y 1970 se desarrolló un énfasis en los objetivos del dominio cognoscitivo y afectivo.

Es común observar en los ambientes educativos que las mediciones que realiza el maestro sobre los rendimientos de los

alumnos no son equiparables a las que llevan a cabo los físicos en su laboratorio. Las mediciones en psicología y educación se quedan frecuentemente en niveles nominales u ordinales, es por eso que antes de hablar del tratamiento estadístico de las calificaciones asignadas supuestamente a las conductas especificadas en los objetivos de las asignaturas, con fines predictivos, sea indispensable aclarar lo que se entiende por medición, niveles de medición, evaluación e importancia de una adecuada redacción de objetivos.

1-2 MEDICION Y EVALUACION EN EL PROCESO EDUCATIVO.

Existen numerosas definiciones de medición, algunos entienden por ella al método por medio del cual se establece una correspondencia única y recíproca de todas o algunas magnitudes de clase y todos o algunos números, sean estos enteros, racionales o reales, según el caso; para otros es la asignación de números para representar propiedades de los sistemas materiales - que no son números, en virtud de las leyes que rigen estas propiedades (5).

Glass y Stanley (6) consideran que: "Medición es la asignación de números a cosas de acuerdo a reglas". En otras palabras: "La medición transforma ciertos atributos de nuestras percepciones en cosas familiares llamadas números". Esta última,

(5) ROMERO, L., 1976. Apuntes de clase de la materia psicometría. México: UAM.

(6) GLASS, V.G. and J.C. STANLEY. 1970. "Statistical Methods in Education and Psychology. New Jersey: Prentice-Hall, Inc. 7-8.

parece ser, la definición que con menos palabras explica la esencia de la medición; ahora bien, existen diferentes grados de medición y es a medida que se avanza a través de las diferentes escalas, que las observaciones y las mediciones se van volviendo más exactas.

1-2.1 Escala nominal.

Se está trabajando en este nivel cuando se describe algo dándole el nombre de uno u otro conjunto particular de categorías o clases mutuamente exclusivas; entendiéndose por mutuamente exclusivas, la imposibilidad de clasificar una sola observación en dos diferentes categorías al mismo tiempo. Las clases en esta escala se refieren básicamente a nombres, aunque en algunas ocasiones se utilicen números para clasificar, mismos que algunos autores suman o restan como cualquier otro numeral; sin embargo, cualquier operación realizada con números a nivel de escala nominal carecerá de significado y no implicará cosa alguna acerca de los objetivos en sí mismos, pues en este tipo de medición cuando se asigna 0 (cero) a mujer y 1 (uno) a hombre, por ejemplo, únicamente se utilizan para distinguirlos y no se desprende de aquí que el hombre tenga mayor "valor" que la mujer.

La escala nominal no es usada prácticamente en medición educativa, ya que no puede proporcionarnos información acerca de que tanto se modificó la conducta de un alumno en relación con los objetivos de la materia.

1-2.2 Escala ordinal.

Este nivel de medición es posible cuando se detecta un cierto orden de los objetos o eventos con respecto a algún atributo, aunque no es factible discernir qué tanto es más grande una cosa que otra, ya que en caso de asignar números, estos indican únicamente el lugar que ocupan en un continuo. Es posible sumar, restar, multiplicar o realizar cualquier tipo de operación aritmética con dichos números; no obstante, puede suceder que los resultados obtenidos después de efectuar cualquier operación no digan cosa alguna acerca de la cantidad de la propiedad que se está midiendo. Como ejemplo de medición a este nivel se tiene: la dureza de los minerales, los números asignados a los concursantes en un certamen de música y la mayoría de los grados de aprovechamiento que asignan los maestros a los rendimientos de los alumnos de su materia. Es aquí evidente que en múltiples ocasiones se realizan diferentes operaciones aritméticas con dichos números; la pregunta es: ¿qué significado tienen los resultados?

1-2.3 Escala de intervalos.

La diferencia sustancial de esta escala con la ordinal es que el tamaño de los intervalos es igual en este caso; es posible distinguir a este nivel de medición diferentes cantidades de un atributo entre varios objetos y establecer además, diferencias iguales entre los objetos.

En este tipo de escala es necesario definir la unidad de medición que podría ser en el caso de la temperatura, el grado centígrado o el grado Fahrenheit. Un hecho importante de ser notad

do es que 0 (cero) en esta medición no significa la falta absoluta del atributo que se mide y es en consecuencia arbitrario: cero grados centígrados no involucra la falta de temperatura, simplemente, por convención, se decidió establecer como cero el punto de fusión del agua, sin significar esto que la energía cinética de las moléculas en este punto fuera nula.

Es así, que además de proporcionar una idea de orden, la diferencia entre los números es significativa; en consecuencia, es posible inclusive traducir grados centígrados a Fahrenheit y viceversa.

La asignación de calificaciones a los aprovechamientos por norma (curva normal), es considerada como medición intervalar por algunos, aunque se discute aún lo referente a si el tamaño de los intervalos es igual o no. De cualquier forma, la asignación de calificaciones por norma, no se encontrará estrictamente a nivel de escala ordinal sino en algún punto entre los niveles ordinal e intervalar.

1-2.4 Escala de razón.

La diferencia fundamental de este nivel de medición y el intervalar, es que el punto cero indica la ausencia total del atributo medido y en consecuencia es posible realizar todo tipo de operaciones con números y estos tendrán significado; así es posible decir que 100 grados Kelvin indican el doble de temperatura que 50 grados Kelvin.

En psicología y educación se trabaja básicamente en

los primeros tres niveles, utilizando en la escala de razón casi exclusivamente mediciones de tiempo.

1-2.5 Medición y evaluación.

Lo que en este trabajo es considerado como proceso de medición, en la asignación de calificaciones en educación, es la construcción en la que se involucra la especificación de objetivos de aprendizaje, la administración y la asignación de puntajes a las pruebas. Evaluación es el proceso dentro del cual quedan comprendidas la interpretación de los puntajes y la toma de decisiones para determinar si son adecuadas o no para ciertos propósitos; es en este último punto en donde para tomar una decisión congruente, antes de que termine el curso, se propone para fines predictivos, la utilización del teorema de Bayes (la asignación de calificaciones como juicio de valor sobre el rendimiento de un alumno involucra de hecho una predicción sobre su comportamiento futuro con respecto a los atributos medidos) como punto medular en este trabajo, ya que este tratamiento estadístico permite tomar decisiones más precisas que el método tradicional de estimar probabilidad, hecho que se discutirá más ampliamente adelante.

Se observa que en la evaluación se emiten juicios de valor sobre la medición de la ejecución y en consecuencia casi siempre existe un mayor grado de subjetividad que durante la medición.

Podría objetarse al uso del teorema de Bayes en evaluación, el que ésta sería subjetiva se se realiza en función

de las expectativas vertidas en los registros de calificaciones personales del maestro; sin embargo, la medición y la evaluación tendrán necesariamente aspectos subjetivos pues dependen de los juicios de maestros y administradores y no sería deseable caer en la trampa de decir que sólo deben emplearse las mediciones "objetivas" o los juicios de los maestros. En un problema tan difícil sería deseable contar con las dos fuentes de información, muy va liosas por cierto.

1-2.6. Problemas de la medición y evaluación.

Dentro de los múltiples problemas que afectan una medición y evaluación de las ejecuciones se tiene una redacción inadecuada de los objetivos de aprendizaje o la completa carencia de ellos, defectos en la construcción de las pruebas, ignorancia por parte de los maestros sobre los factores que disminuyen la confiabilidad y la falta de un tratamiento estadístico a los puntajes obtenidos en las pruebas.

La predicción de calificaciones bayesiana, permitirá, como se espera demostrar, que dada su mayor exactitud en la predicción, las decisiones escolares mejoren, pudiéndose facilitar en mayor grado la clasificación y colocación, la selección y el diagnóstico de aptitudes y habilidades especiales.

Es un hecho que los registros de calificaciones de los alumnos son utilizados para realizar predicciones sobre el éxito probable de los mismos en el desempeño de diferentes actividades, asimismo, para obtener una beca, un empleo o el ingreso a un centro de estudios superiores, se toma en cuenta el promedio de ca-

lificaciones de los alumnos. El carácter predictivo de la evaluación se torna evidente y siendo el sistema bayesiano un procedimiento estadístico para la realización de inferencias, y habiéndose probado ampliamente su eficacia en diferentes campos, no existe razón de peso para no utilizarlo en el campo educativo.

1.3 LA IMPORTANCIA DE LA CORRECTA ESPECIFICACION DE OBJETIVOS EN LA EVALUACION.

¿Qué es lo que se pretende medir y posteriormente evaluar? ¿Con base en qué se emitirán juicios de valor sobre lo adecuado o no de ciertas ejecuciones con respecto a propósitos específicos? Y algo más: ¿cuáles serán dichas ejecuciones? ¿En qué condiciones realizarán los alumnos las conductas esperadas? ¿Cuáles serán los criterios de ejecución que serán considerados como aceptables para el logro de los objetivos? Presumiblemente, lo que pretenden medir las pruebas de rendimiento son los aprendizajes de los alumnos sobre diversos tópicos; tomando en cuenta que el aprendizaje tiene el carácter de una variable intercurrente, resulta complicado medirlo directamente, limitándose el proceso de medición, a las manifestaciones del mismo. Y son los cambios de conducta, resultado de la experiencia, lo que se tratará de medir.

Hilgard y Marquis (7) definen fácticamente al aprendizaje como: "Un cambio relativamente permanente en la potenciali-

(7)HILGARD, E.R., y D.G.MARQUIS.1972. "Condicionamiento y Aprendizaje".México: Trillas. Pág. 17.

dad de la conducta que ocurre como resultado de la práctica reforzada". Tomando en cuenta que el aprendizaje puede manifestarse como un cambio de conducta, el papel del maestro sería precisamente modificar el comportamiento de los alumnos, mismo que sería medido con pruebas de rendimiento con fines de evaluación diagnóstica, formativa o sumativa.

"Un objetivo es el fin o la meta a que se dirige una acción u operación, es el término que persigue un proceso o una de sus fases" (8). Algunos autores diferencian "meta" de objetivo", aquí serán considerados como sinónimos. Ahora bien, un objetivo educativo o de aprendizaje es: la descripción clara y unívoca de lo que debe hacer el alumno al terminar el ciclo escolar". "Un objetivo (educativo) es un propósito expresado en un enunciado que describe el cambio propuesto en el alumno cuando haya terminado exitosamente una experiencia de aprendizaje" (8)

Una vez que el maestro cuenta con los objetivos de aprendizaje de su materia y con aquéllos que describen las conductas de los requerimientos básicos para las mismas, es factible de acuerdo a la estrategia de enseñanza presentada por Anderson y Faust (9), realizar una premedición, con base en la cual se realizará una evaluación diagnóstica, en la que se determinará si los conocimientos previos son suficientes y si los alumnos ya poseen o no los objetivos propuestos para tomar en con-

(8) MAGER, R. 1969. "La confección de Objetivos para la Enseñanza". Cuba. Año de los diez millones. Pág. 17.

(9) ANDERSON, R.C. y G.W. FAUST. 1977. "Psicología Educativa". México: Trillas. Págs. 147-195.

secuencia las decisiones que se consideren pertinentes. En el caso de que los alumnos ya posean los conocimientos previos indispensables y no dominen las conductas expuestas en los objetivos del curso, se procederá a la elaboración o selección de los métodos de enseñanza, a llevar a cabo ésta, a medir nuevamente y a evaluar.

Tanto en la evaluación realizada con el pretest como en la sumativa del último paso del proceso, se toman decisiones de relevancia. Considerando esta estrategia de enseñanza como un curso unitario o bien como una unidad de programa, es viable medir ya sea total o parcialmente los logros de los objetivos propuestos.

Si bien, se ha tratado la importancia de una redacción de objetivos clara y unívoca, aún no se han discutido las características esenciales con que debe contar un objetivo para resultar comunicador, útil para incrementar los rendimientos académicos y sobre todo, ya que esto es lo que interesa actualmente, práctico para implementar la construcción de instrumentos de medición, propiciar una evaluación acertada de las ejecuciones y una predicción de notas (rendimientos) que involucre un significado relacionado con la medición de las mismas.

Existe una variedad enorme de posiciones en lo referente a la redacción de objetivos de aprendizaje conductuales, las hay que suponen que con añadir un verbo de categoría "observable" al sujeto que desempeñará la acción es suficiente y existen otras que postulan como deseables hasta quince requerimientos diferentes.

Remontándose el estudio a la casi prehistoria de los objetivos, se encuentra la postura de Robert F. Mager en la obra anteriormente citada, sobre redacción de objetivos, que propone las siguientes cualidades para la confección de objetivos de aprendizaje:

Conducta observable: entendiendo por conducta a toda actividad susceptible de ser observada. Mager propone la elección de un verbo activo que en la medida de lo posible sea sujeto a pocas interpretaciones.

Condiciones: son básicamente las ayudas y restricciones con que contará el alumno cuando se le pida demostrar el logro del objetivo realizando la conducta pedida.

Criterio de ejecución aceptable: si las condiciones establecen **algunos** aspectos de importancia fundamental en los resultados de las mediciones, el criterio de ejecución aceptable (entendido como la descripción en términos de porcentajes, índices o palabras que indiquen la calidad de la ejecución; o sea, de las características que debe tener la conducta del alumno para ser considerada como satisfactoria) resulta de importancia fundamental para la medición educativa.

A estos criterios, sería deseable agregar el de evitar palabras innecesarias para no volver engorrosa e impráctica la lectura de los objetivos.

La determinación de si el objetivo es muy importante, de regular importancia o complementario, resulta útil al planear un balance adecuado de reactivos en los instrumentos de medición. La clasificación taxonómica entera del Dominio Cognoscitivo de Bloom y colaboradores también sería deseable, para mejorar la

comunicación y cuidar el nivel de dificultad de las preguntas de examen que supuestamente deben concordar con lo expresado en los objetivos. La clasificación "entera" que se menciona anteriormente hace referencia a las categorías de conocimiento, comprensión, aplicación, análisis, síntesis y evaluación; sin llegar a sofisticaciones tales como "conocimiento de tendencias y secuencias", ya que esto dificulta la clasificación y no agrega gran cosa a la eficiencia de los procedimientos de medición y evaluación.

Es indispensable entonces, una redacción de objetivos clara y no susceptible a múltiples interpretaciones; las categorías recomendadas deberán incluirse siempre y cuando mejoren la comunicación. Si se sabe lo que se pretende de los alumnos, además de que todos los esfuerzos se encaminarán hacia esa meta, la medición realizada será congruente con lo que presumiblemente se enseñó, la evaluación basada en el tratamiento estadístico de los puntajes en bruto tendrá algún significado, pues se sustentará sobre la base de mediciones de las ejecuciones menos subjetivas y la predicción de calificaciones por el método de Bayes que se propone, será de utilidad.

1-4 FORMULACION DEL PROBLEMA DE ASIGNACION DE CALIFICACIONES.

Una vez que se han utilizado diversos procedimientos para realizar las mediciones de las ejecuciones de los alumnos -con respecto a los objetivos del curso o unidad- se torna difícil el traducir la calidad de los rendimientos de los estudiantes, realizadas a lo largo del ciclo escolar, en notas o califi-

caciones. Se ha medido y registrado la ejecución del alumno, lo que debe realizarse en consecuencia es: la comunicación de la evaluación - realizada por el maestro- de los resultados de la medición. Sin embargo, sucede que en la mayoría de los sistemas escolares se llevan a cabo procedimientos para otorgar notas utilizando números o letras. Se resume en uno o dos símbolos abstrac-tos el desempeño de un alumno durante 6 meses o 1 año, en un símbolo que no dice gran cosa acerca de qué objetivos de la materia se lograron y de si aquéllos conseguidos fueron los más o los menos importantes.

Aunado a este hecho, se tiene el de la conversión de los puntajes de las ejecuciones, que suele ser arbitraria y ba-sada en múltiples ocasiones en juicios de valor insuficientemen-te justipreciados de profesores y administradores.

La dificultad intrínseca de las mediciones sobre las actuaciones de los alumnos, debida en parte a la ausencia o re-dacción inadecuada de objetivos de enseñanza-aprendizaje conduc-tuales, a la inconstancia del objeto de medición y a la comple-jidad de los procedimientos de medición, es otro aspecto a con-siderar.

Además, los diferentes sistemas de asignación de cali-ficaciones no cuentan con un patrón común de aplicación e inter-pretación, por lo que los diferentes maestros actúan según aquello que consideran más pertinente. El resultado es un caos absoluto, pues cada profesor aplica su criterio, teniendo así, que existe una multiplicidad de significados con respecto a lo que

una nota o calificación revela. Significa un 10, una A, un MB o una S que la ejecución de Juan Pérez comparándola con la de sus compañeros de clase fue la mejor? ¿Significa que se esforzó mucho a juicio del maestro aunque su ejecución apenas alcanzó a cubrir el 50 % de los objetivos de la materia? Significa que logró un completo dominio de los objetivos de la asignatura?

El problema de la asignación de notas o calificaciones ha sido ampliamente discutido: Johnson en 1911 (10) pone de manifiesto en su trabajo que la asignación de puntos a las pruebas y la asignación de notas, previas a la aparición de las pruebas objetivas, eran exactamente lo mismo.

El estudio de Starch en 1913 (11) dio a conocer la gran subjetividad que existía por parte de los maestros en la asignación de calificaciones. Dadourian en 1925 (12) propuso la abolición de las calificaciones basado en la argumentación de - que producían serios daños en la personalidad de los niños cuando eran calificados de manera inadecuada.

Ebel en 1965 (13) encontró que en las diferentes ma-

(10) JOHNSON, F.W., 1911. "A study of high school grades". School Review 19. Págs 13-24.

(11) STARCH, D., 1913. "Reliability and distribution of grades. Science, 38. Págs 630-636.

(12) DADOURIAN, H.M., 1925. "Are examinations worth the price? School and Society, 21. Págs 442-443.

(13) EBEL, R., 1965, "Measuring educational achievement". - Englewood Cliffs, N.Y.: Prentice Hall.

terias y departamentos de una misma escuela, existía cierta tendencia a asignar notas A o B, mientras que un departamento emitía 63 % de calificaciones A y B, en otro el porcentaje era de 26. Asimismo, afirma que los criterios para asignar calificaciones altas (A y B) variaban con referencia al tiempo, observando que en un lapso de 18 años se había observado un incremento constante del 33 al 44 % en una escuela de Estados Unidos.

Glasser en 1969 (14) elabora una argumentación en contra del uso de letras o números como calificaciones. Por su parte Moynihan en 1971 (15) afirma que aunque los problemas de la asignación de calificaciones son de gran importancia, la incorrecta información que proporcionan es mejor que no contar con ninguna vía de realimentación; dice también, que uno de los grandes logros de la "democracia" es el sistema de calificaciones que permite que las personas talentosas menos favorecidas económicamente puedan ser reconocidas por su valor. Opina que el sistema puede parecer cruel o aristocrático pero que realmente es lo contrario, pues en un mundo en donde no existiera, únicamente el in-fluyentismo operaría.

Robert F.Mager en 1975 (16) afirma que a menos que se den a conocer públicamente todos los objetivos que el "aprobado" o la calificación numérica o por letras representan, el cam-

(14)GLASSER, W.,1969. "Schools Without failure". New York:Harper and Row,Publishers.

(15)MOYNIHAN, P.,1971. "Seek parity of educational achievement" Moynihan urges. Report on Educational Research (march 3), 3(5).p.4

(16)MAGER,R.1975."Medición del Intento Educativo".B.A.:Spe.Pág 105

bio del sistema de notas por letras o números al de "aprobado", "reprobado", no aporta ningún beneficio.

A pesar de los múltiples problemas que involucra la asignación de notas, las funciones que desempeñan son de vital importancia en los campos individual, social y económico de un país; por lo que siendo el tema central de este estudio la aplicación de un método estadístico a las calificaciones iniciales, con el fin de predecir la calificación final, se procederá a realizar una revisión de las principales funciones que desempeñan las calificaciones.

Funciones administrativas: las calificaciones como medios de acreditación y certificación, cuando han sido establecidos estándares de ejecución, permiten la demostración de dominio en ciertas áreas o profesiones, por lo cual es posible, emitir los certificados de terminación de estudios a las instituciones correspondientes. Para la toma de decisiones, las calificaciones pueden ser una vía para diagnosticar a aquellos individuos que necesitan algún tipo de enseñanza remedial, o para seleccionar a aquéllos que serían candidatos ideales para un curso especial de matemáticas, o para recibir una beca de estudios. El método que se propone en este trabajo pretende ser de utilidad precisamente en el cumplimiento de esta función: la toma de decisiones. Con referencia a la clasificación y agrupamiento de los alumnos de acuerdo a los diferentes niveles de habilidad demostrados, en el caso de que esto se considere adecuado, las notas escolares proporcionarían un medio para colocar a los alumnos en grupos idóneos.

Las calificaciones también son utilizadas para decidir que alumnos acreditarán o repetirán el curso. De la misma manera, contribuyen a la **decisión** de que el alumno no permanecerá en la escuela debido al número de materias reprobadas en su registro. Las calificaciones son un medio de control de calidad cuando son utilizadas con referencia a normas locales o nacionales, o a pruebas referidas a un criterio, proporcionan un medio de detección de la calidad de la enseñanza en una escuela.

Funciones de enseñanza: realimentación al maestro, dada la construcción de un sistema adecuado de medición, realizado - con base en los objetivos del curso, las calificaciones que emergieran de ella, proporcionarían información útil al maestro sobre qué unidades no fueron adecuadamente cubiertas o presentan mayor dificultad para los alumnos. Del mismo modo, el alumno puede identificar áreas de debilidad específicas y ambos pueden en consecuencia tomar decisiones encaminadas a solucionar los problemas identificados.

Realimentación al alumno y a sus padres o tutores: tanto uno como los otros, requieren información sobre qué objetivos han sido alcanzados y cuáles no; los estudiantes deben saber qué es lo que necesitan aprender.

Para llevar a cabo las funciones de enseñanza, es evidente que una letra, un número o un "aprobado" no cumplen ampliamente con los requerimientos.

Si bien H.B. White en 1932 (17) encontró que la ejecución de los alumnos tiende a mejorar por el solo hecho de realizar un mayor número de exámenes, las calificaciones en sí mismas no proporcionan una mejora en la calidad de las ejecuciones.

Funciones sociales: tanto al alumno como al orientador vocacional, las calificaciones escolares proporcionan, a pesar de las deficiencias de las mismas, indicios sobre las áreas específicas de conocimientos en las que los desempeños probables serán de mayor calidad.

Las calificaciones influyen en la selección de los supuestamente mejores individuos para ir funcionando como una especie de colador que siguiendo las leyes de la selección natural, permita el paso de los más aptos a los niveles de educación media y superior. El ingreso a los centros de educación superior se basa en parte en las calificaciones obtenidas en los ciclos de enseñanza básica y media.

Cuando por alguna razón económica-política es necesario reducir el número de egresados de alguna profesión o ciclo de enseñanza, una forma de lograrlo es incrementar la dificultad de los exámenes para que las notas asignadas a los desempeños de los alumnos sean en gran proporción bajas.

La decisión de reprobar o no a un alumno o de expulsarlo de la escuela, basada en parte en el número de materias reprobadas tiene repercusiones sociales y económicas de interés para

WHITE, H.B. 1932. "Testing as an aid to learning". Educational Administration and Supervision, 19. Págs 41-46

el individuo y para la comunidad. Aquí valdría la pena cuestionar, si la función de la evaluación es de carácter selectivo, para permitir la entrada de los individuos más aptos a los niveles de vida de mayor ingreso o si la evaluación debe tener el carácter de desarrollo, entendido éste como el facilitar que la mayoría, si no todos los individuos integrantes de la sociedad, alcancen el más alto nivel de educación posible. Se tiene el caso de la Universidad de Pardue, Indiana, en la que la función de la evaluación es básicamente formativa, en el sentido de detectar los defectos de los medios y métodos de instrucción, así como las áreas de debilidad de los alumnos para instrumentar cursos de nivelación y actualización. Algunos alumnos terminan sus profesiones en lapsos que van desde 2 a 7 años, dependiendo de las habilidades de los mismos y de los conocimientos previos con que ingresan en la escuela.

Claro está, que una política de esta clase implica una erogación monetaria muy superior a la permitida por los precarios presupuestos asignados a las escuelas en países subdesarrollados y en crisis de "liquidez" como México.

Tratamiento estadístico de los puntajes:

La estadística descriptiva es aquel conjunto de técnicas que involucra la tabulación, representación gráfica y descripción de colecciones de datos describiéndolos o resumiéndolos a formas más manejables. La estadística inferencial, es un cuerpo estructurado de técnicas que tratan de inferir las propiedades de un conjunto muy grande de datos a través de la inspección de una muestra del conjunto, obteniendo inferencias de propieda-

dos particulares de las muestras y aplicándolas a propiedades particulares de las poblaciones. (18)

Sin embargo, la evaluación se encuentra en la desafortunada posición de tener que depender de mediciones indirectas ya que a la fecha no se ha encontrado algún medio de medir directamente el éxito de un alumno. Pareciera aquí que si no se ataca el problema de la medición tratando de obtener medidas - más confiables y más válidas, el uso de una técnica estadística u otra cualquiera no sería de gran utilidad. Si un maestro no elabora sus exámenes en función de objetivos conductuales - claramente redactados y sigue los cánones encontrados por diversos investigadores para incrementar la confiabilidad y validez de los exámenes, no importa que tan sofisticada manipulación estadística utilice para el manejo de sus datos, **los resultados** - que proporcionen los métodos no ofrecerán mayor significado que el de la calidad de los datos de las mediciones que ha manejado.

El manejo estadístico de los datos sólo podrá proporcionar resultados válidos cuando los procesos que subyacen a la evaluación educativa han sido cuidadosamente realizados.

Número de categorías a utilizar:

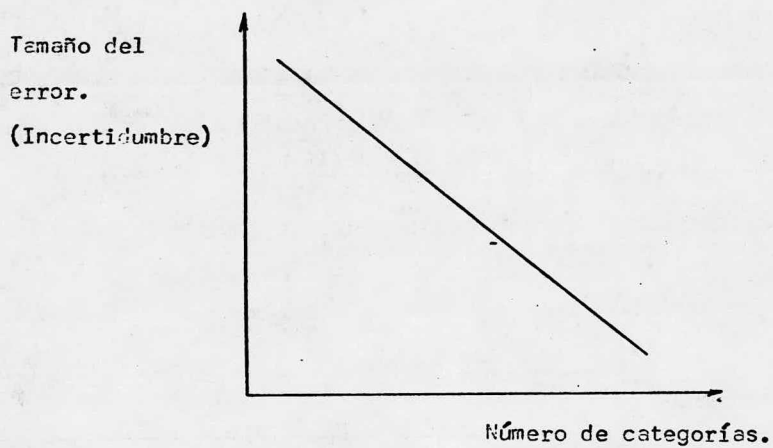
Independientemente de si se utilizan números o letras, el problema estriba en la determinación del número más apropiado de categorías que los profesores puedan utilizar con eficacia y que proporcione, simultáneamente, la mejor información posible.

Por un lado se tiene el hecho de que en los diferen-

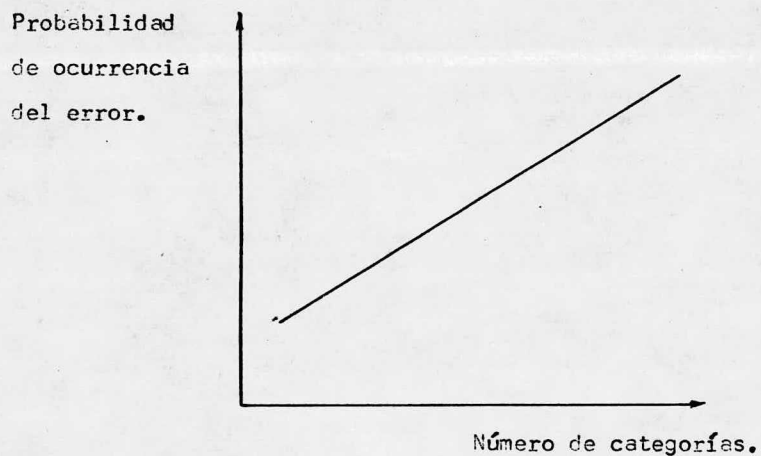
(18) GLASS, V.G., and J.C. STANLEY. Op. cit. Pág.2.

tes ciclos escolares y en las diferentes instituciones educativas se utilizan números variables de categorías. En algunos casos las evaluaciones se expresan utilizando únicamente dos categorías: aprobado o reprobado; en otros tres o cuatro, como en la UAM: MB (muy bien), B (bien), S (suficiente), NA (no acreditado); once en algunas escuelas dependientes del IPN: de 0 a 10; cinco letras en algunas universidades particulares del país; - quince, como sucede en algunos distritos de Estados Unidos en - los que agregando los signos más (+) o menos (-) a las cinco categorías de mayor uso: A, B, C, D, y F logran un sistema de 15 categorías a utilizar; ó 101 categorías en el sistema de porcentajes que aún se encuentra en uso en algunas escuelas.

Por un lado, existe una relación directamente proporcional entre el número de categorías y la cantidad de información que proporcionan; así, si sólo se cuenta con un "aprobado" o "reprobado", la información que este nivel ordinal proporciona es bastante escasa. De la misma manera, el error que se puede cometer al dar un "reprobado" cuando el alumno ha cubierto - los objetivos de la materia satisfactoriamente, o de dar un "aprobado" cuando no los ha cumplido, será de consecuencias importantes. Se tiene también, que al aumentar el número de categorías, se reduce el tamaño del error a costa de incrementar las - probabilidades de ocurrencia del mismo y viceversa. Lo anteriormente expuesto puede ilustrarse mediante la siguiente representación gráfica, en la que con propósitos de simplificación se - ha supuesto una relación lineal entre las variables.



(a)



(b)

Figura 1. (a) Tamaño del error vs. número de categorías. (b) Probabilidad de ocurrencia del error vs. número de categorías.

Por lo anteriormente expuesto, parece ser que un número limitado de categorías (por ejemplo dos) involucraría un tamaño mayor del error; sobre este particular, Thorndike y Hagen (19) afirman que el hecho fundamental en el cual debe basarse el número de categorías de calificación, es la confiabilidad de los instrumentos de medición en que se basará la decisión.

Los resultados de estos autores se resumen en la tabla 1.

Tabla 1
Porcentaje de calificaciones. Errores.

Confiabilidad No. de categorías. Prueba	0.6	0.7	0.8	0.9
15 (A a F-)	78	69	63	50
5 (A a F)	30	30	15	4
3 (Sobresaliente, aprobado, reprobado).	9	5	2	0.01

Sin embargo, la variación de por lo menos 3 puntos en un tercio de las calificaciones y de como mínimo un punto en el 37 % de los casos, cuando se tienen ciento una categorías, no es un error de graves consecuencias (el tamaño del error es relativo).

(19) THORNDIKE R., y E. HAGEN. 1970. "Tests y técnicas de medición en psicología y educación". México: Ed. Trillas. 601-602.

vamente pequeño). Los supuestos en que se basa el cuadro anterior no han sido suficientemente explicados.

Los resultados de Peters y Van Woorhis pueden representarse según la tabla No. 2. (20)

Tabla 2

Tamaño del error (no confiabilidad de las notas).%

Confiabilidad Prueba No. de categorías.	0.50	0.70	0.80	0.90	0.95
2	67	53	47	40	37
5	55	38	29	20	15
10	52	32	22	13	8
15	51	31	21	11	6

Según esta tabla, si las calificaciones están basadas en una información que posee una confiabilidad de 0.95, el uso de dos categorías en las mismas implicará un error de 37 %; con 5 categorías, de 15 % ; y empleando 10 categorías, el error sería de escasa importancia .

La información proporcionada por las dos tablas anteriores resulta complementaria; sin embargo, la información de la primera no es lo suficientemente explícita, pues no explica la fundamentación de los datos. Ebel en 1965 (21) coincide con (20) PETERS, F.R. 1956. "Measurement of Informal Educational Achievement". School Review, 64: 227-232.

(21) EBEL, R.L., 1965. "Measuring educational achievement". - Englewood Cliffs, N.J.: Prentice-Hall, Inc. citado en STANLEY J. and K. HOPKIN. 1976. "Educational and psychological measurement and evaluation. Englewood Cliffs, N.J.: Prentice Hall, Inc.

los resultados del cuadro de Peters por lo cual podría concluirse, dada la información con que se cuenta, que con menos categorías se pierde una gran cantidad de información acerca de las diferencias individuales dentro de cada categoría. Existe un menor número de personas en las que se cometen errores, pero la gravedad de los mismos se incrementa de manera considerable, sobre todo si sólo se cuenta con dos categorías. Actualmente en la UNAM se cuenta con un sistema basado estrictamente en cuatro categorías, mismo que fue cambiado de uno de once, decrementándose de esta manera la confiabilidad de las calificaciones.

1-4.1 Significado de las Calificaciones:

Existe una gran variedad de puntos de vista acerca de qué aspectos deben ser considerados para la asignación de calificaciones. Algunos maestros toman en cuenta la cantidad de trabajo llevada a cabo por los alumnos; es frecuente observar que se solicita a los alumnos con el fin de dar una mejor nota la realización de algún trabajo extra. Así, la calificación asignada por el profesor significará, tal vez, el grado de resistencia a la extinción del alumno, entre otras cosas.

En algunas ocasiones, la calificación es asignada en función de la claridad de la exposición oral o escrita, de la buena letra o de la presentación y claridad del trabajo o examen; en este último caso, la calificación tendría un significado desligado de los objetivos de la materia- a menos que sea redacción-. En suma, un gran número de maestros utiliza además de los puntajes obtenidos a partir de las mediciones un sinnúmero de aspectos subjetivos, tales como puntualidad, esfuerzo, claridad del

trabajo escrito, el "efecto de aura" producido por el prestigio ganado por el alumno con base en calificaciones o actuaciones a-priori. Así se encuentra que en los ciclos escolares elementales existe una tendencia a otorgar mejores calificaciones a aquellos alumnos que se encuentran realizando su mejor esfuerzo, no importando que la calidad de las ejecuciones no sea ni la esperada, ni de las mejores del grupo; de la misma forma se ha observado que se asignan notas más bajas a aquellos alumnos que no se encuentran realizando, a juicio del maestro, un esfuerzo equiparable a su capacidad, aunque la calidad de sus ejecuciones sea mejor que la de los alumnos del primer grupo.

El significado de las calificaciones se manifiesta vago cuando se mezclan diferentes aspectos conductuales de los alumnos, con la medición de las ejecuciones de los estudiantes. Sin embargo, el fin para el cual va a ser utilizada una calificación, debería ser la pauta que indicara en qué aspectos deba basarse la asignación de notas.

1-4.2 Marco de referencia para la asignación de calificaciones.

Según Thorndike y Hagen (22) existen tres marcos de referencia posibles en relación con los cuales se asignan las calificaciones:

Ejecución en relación con la perfección.

Ejecución en relación con el potencial.

Ejecución en relación con los iguales.

1-4.21 Ejecución en relación con la perfección: El primer marco de referencia se relaciona con el grado de dominio que demuestre un sujeto en relación con los objetivos de la materia cuando

(22) THORNDIKE. Op cit. Pág 604.

estos han sido especificados conductualmente; siendo posible así, realizar un examen en el que se incluyan todos los objetivos de una unidad o curso o una muestra representativa de los mismos. Una forma posible para asignar la calificación es reportarla en forma de porcentaje de la ejecución perfecta. Algunos autores afirman que el número de casos en que se puede hablar de ejecución perfecta es pequeño y que en consecuencia este marco de referencia tiene una utilidad muy limitada.

1-4.22 Ejecución en relación con el potencial: este segundo marco de referencia asigna las calificaciones en función de alguna estimación del potencial de ejecución. Argumentando en favor de la utilización de este medio de comparación se habla de la utilidad e importancia de la información, de si la ejecución de un alumno está en función de su habilidad potencial o nivel intelectual (cualquier cosa que esto sea). Sin embargo, en algunas situaciones es necesario decidir si un alumno debe inscribirse a un curso más avanzado o abandonar la escuela; en este caso es importante saber qué puede realizar ese alumno en particular, fundamentalmente.

1-4.23 Ejecución en relación con los iguales: este último marco de referencia realiza la comparación de las ejecuciones de un individuo con algún grupo de referencia; es el que se utiliza en el sistema de asignación de calificaciones con respecto a la norma, también es el que se utiliza en las pruebas estandarizadas, en las cuales el grupo de referencia trata de ser una muestra representativa de una población. Una pregunta de interés sería: ¿con cuál grupo de iguales se realizaría la com-

paración? ¿Con los miembros de su clase? ¿Con el formado por todos los alumnos de la escuela que toman esa materia? ¿Con el formado por todas las escuelas del distrito, del estado o del país? ¿Cuál de todos los grupos anteriores proporcionarán el mejor marco de referencia para este grupo particular? Tal vez las condiciones de enseñanza, el nivel socioeconómico de los alumnos y las características de la escuela determinen la comparación más adecuada de acuerdo a este marco de referencia con los iguales y quizás una comparación con un grupo nacional no sea la más apropiada.

Existen ciertas argumentaciones en favor de la utilización como marco de referencia para la asignación de calificaciones de un grupo en particular, la comparación con las del grupo formado por la totalidad de los alumnos de esa materia, en ese nivel, en el país; dentro de estas razones, se encuentra la de que a un alumno debe dársele la oportunidad de obtener determinada nota, si la calidad de sus ejecuciones es semejante a la de otro estudiante; sin tomar en cuenta: el nivel socioeconómico, la escuela, los procedimientos de enseñanza aprendizaje, los maestros y la región del país en la que se encuentra; otro argumento, íntimamente ligado al anterior, es el de que de esta manera, si los objetivos e instrumentos de medición son uniformes, el significado de una calificación asignada por un maestro en una escuela determinada será semejante a la asignada por otro profesor en otro lugar; así, la comparación entre los individuos se volvería más significativa. No obstante, los dos argumentos enunciados anteriormente son refutados cuando se dice que, para

un alumno en particular, es importante la posibilidad de obtener una calificación alta en su propio contexto económico, social y cultural; para lo cual debería utilizarse como marco de referencia ya sea su propia clase o el conjunto formado por los grupos de esa materia en el mismo nivel dentro de la escuela.

También se afirma, que el marco de referencia adecuado sería el estructurado en función de la "potencialidad" de ese alumno en particular, o sea evaluar y asignar notas en términos del progreso del estudiante con base en el nivel previo de competencia del mismo y no en relación con las ejecuciones de sus compañeros, más bien utilizando como marco de referencia su potencial de realización.

Cuando se utiliza un grupo específico como marco de referencia para otros grupos, podría suceder que los alumnos integrantes del conjunto no fueran representativos de la población y que en consecuencia las notas resultantes no fueran adecuadas, pues algunos alumnos recibirían calificaciones altas o bajas (según el caso) en función con las características del grupo de referencia.

El marco de referencia que se utilizó para la asignación de notas en este estudio: calificaciones de Fisicoquímica en la Escuela Superior de Ingeniería Química e Industrias Extractivas del IPN fue el de ejecución en relación con la perfección. Los objetivos de la materia fueron específicos de acuerdo con los lineamientos anteriormente mencionados y los reactivos de examen cuidadosamente elaborados y muestreados para la confección de los exámenes. Las calificaciones fueron asignadas de acuerdo con el grado de dominio demostrado por el alumno en re-

lación con los objetivos de la materia. Sin embargo, la calificación no se reportó en forma de porcentaje de la ejecución perfecta, sino que se realizó una transformación de los mismos de acuerdo a lo siguiente: de 0-59% -NA, de 60 a 79% -S, de 80 a 94% B, de 95 a 100% -MB.

II.-EL METODO BAYESIANO.

2.1 Interpretación de probabilidades.

La medición de la incertidumbre requiere de un lenguaje que describa los resultados, este lenguaje lo proporciona la teoría de la probabilidad. Si se desea comprender qué significa que la probabilidad de lluvia sea del 90%, por ejemplo, es necesario penetrar en la terminología de la probabilidad y explicar los principios y las reglas para realizar cálculos sobre ellas, por lo que a continuación se tratarán de explicar aquellos conceptos sobre probabilidad que resultan indispensables para la comprensión de la estadística bayesiana.

Cuando se habla de probabilidad, en realidad se intenta medir la incertidumbre y cambios en la incertidumbre; para realizar esto, es indispensable aplicar una medida de incertidumbre -la probabilidad- a aquéllo que interesa.

Para lograr una decisión con respecto a un problema específico no es posible basarse en frecuencias a largo plazo o en lo que sucederá después de muchos ensayos, cuando el interés se centra en lo que va a suceder con una droga específica, con una persona y no con otra, o con el rendimiento futuro de un alumno o un grupo de ellos y siendo que el principal interés es realizar inferencias acerca de estos problemas únicos, y que la base para la inferencia es el conocimiento, se espera demostrar que es el enfoque bayesiano una respuesta viable que permite aplicar "nuestro" conocimiento a la solución de los problemas únicos que ocupan, ahora, la atención del maestro.

Siendo que la probabilidad expresa el grado de creencia o conocimiento sobre algo y que depende de los datos con los que se cuenta y de las suposiciones que realiza el que predice, se justifica la utilización de este enfoque para el problema que se trata: la predicción de los rendimientos de los alumnos con fines de evaluación formativa. Probablemente resultará de suma utilidad para el profesor que desee tomar medidas correctivas, el conocimiento de la calificación final más probable de cada uno de sus alumnos, dada la calificación inicial.

No siempre es posible asignar probabilidades con gran exactitud; sin embargo, se pueden obtener valores razonables por medio de diferentes procedimientos. Permítase considerar el siguiente ejemplo: ¿Cuál es la probabilidad de que Pedro Sánchez obtenga calificaciones altas en matemáticas dado un pretest en el que obtuvo S? En primer lugar, ¿qué se considera calificación alta en matemáticas? Habría que definir esto más precisamente; una respuesta posible al intento de dar un significado menos confuso sería: obtener una calificación de MB. Ahora bien, ¿qué tipo de sistema de evaluación se está usando? ¿evaluación referida a un criterio o al sentimiento del profesor? ¿Qué consistencia interna tienen las mediciones realizadas? Es decir, ¿qué tan "bien" o "mal" se han construido las pruebas aplicadas?, ¿qué grado de concordancia tienen las preguntas del examen con los objetivos de la materia, con el nivel de dificultad de los mismos? ¿Y qué de la confiabilidad de las calificaciones? La predicción que se realice utilizando cualquier método estadístico, tendrá un grado de exactitud función de los datos que han sido proporcionados. Se obtiene en proporción con lo que se da.

Es posible utilizar una gran cantidad de información para estimar las probabilidades: registros anteriores de calificaciones, consistencia de las mediciones, intuición; pero en aquellos casos en que no se cuente con ninguna información, se asignará la probabilidad de $1/2$, con la condición de que el grado de creencia, esto es la probabilidad, cambiará a la luz de nueva información.

Todos los procedimientos estadísticos a utilizar dependen en última instancia de un juicio personal, se cuestiona si el modelo que se pretende utilizar es adecuado a las características que se observan en el problema o no; si la información recolectada es relevante para la predicción que se pretende realizar; si la extrapolación efectuada resulta razonable de acuerdo con lo que se cuenta, en suma: el procedimiento estadístico "más objetivo" está plagado de subjetividad.

Cuando se pretende saber algo de una población, generalmente se reúnen, organizan, describen y se realizan operaciones con los datos, para inferir algunas propiedades de la población. Los datos de la muestra usualmente son utilizados para revisar un conocimiento anterior del que toma decisiones, para generalizar acerca de parámetros de población desconocidos o para extrapolar acerca del comportamiento de una variable o de un conjunto de ellas. Los datos obtenidos de una muestra de la población reducen la incertidumbre sobre los estados de la naturaleza, pero no la eliminan.

La estadística, como un método de toma de decisiones ante la incertidumbre, se basa en la teoría de la probabilidad, ya que la probabilidad es la medida y manera de expresar la incertidumbre y los riesgos asociados a ella.

Previo a la explicación de los procedimientos de estadística bayesiana para la predicción de calificaciones con fines de evaluación formativa, se revizarán algunos conceptos fundamentales sobre probabilidad con objeto de establecer un marco teórico completo para el tratamiento de los puntajes que se propone en este trabajo. Aunque también son necesarios algunos conceptos sobre teoría de conjuntos y análisis combinatorio, estos no serán incluidos en el cuerpo de la tesis, para evitar el aburrimiento de aquéllos que ya conocen éstos tópicos. Dichos temas serán tratados sumariamente sin embargo, en el glosario de términos incluido al final.

2.11 Principios de probabilidad.

La teoría de la probabilidad fue desarrollada en sus inicios por Pascal (1623-1662) y Fermat (1601-1665) para calcular las probabilidades en los juegos de azar.

Cabe ahora realizar una diferenciación entre la teoría de la probabilidad y el cálculo de probabilidades. ¿Qué es probabilidad? y como se interpreta un resultado de 0.6, probabilidad de lluvia, por ejemplo, son asuntos relacionados con la primer cuestión y existen tres escuelas de pensamiento que tratan de dar respuesta a estas cuestiones. En cambio, las deducciones matemáticas obtenidas a partir de los axiomas básicos de probabilidad, son las materias relacionadas con el cálculo de probabilidades.

¿Qué es probabilidad? ¿Qué interpretación se puede dar a la probabilidad? Se sabe que si es seguro que un hecho ocurra, la probabilidad asociada, por convención, será uno y si es imposible que suceda, será cero. Ahora bien, ¿es totalmente posible que un hecho pase? ¿es "totalmente seguro" que algo pasará o que no pasará?

Se observa disentiimiento en la opinión de los diferentes estadísticos, por lo que se encuentran básicamente tres escuelas de pensamiento: la teoría clásica, la teoría de la frecuencia relativa y la teoría personalística o subjetiva; dentro de la que se encuentra la estadística bayesiana.

2.1.2 La teoría clásica. Aunque existe quién clasifique a la teoría clásica dentro de la subjetiva (1); parece ser que esta teoría surgió con los matemáticos franceses del siglo XVII como un intento de dar respuesta a interrogantes de interés pecunario, tales como qué podía considerarse una buena apuesta y si era posible cuantificar la probabilidad de ganar en los dados, la ruleta o las cartas; esta teoría se encuentra basada en el principio de razón insuficiente enunciado por el matemático suizo Jacques Bernoulli (1654-1705) que dice que en un experimento al azar los resultados son igualmente probables: si hay dos posibles resultados, tales como águila o sol; así, si una moneda equilibrada es lanzada, los resultados posibles tienen los mismos pesos y se puede considerar que existe la misma probabilidad de que caiga águila o sol; por lo tanto, la probabilidad de que sea águila o sol será de $1/2$.

Generalizando lo anterior, se tiene que si hay N resultados posibles igualmente probables de un experimento al azar, la probabilidad para cada punto de muestra debe ser $1 / N$. Así:

(1) YAMANE T. (1973), Estadística, México: Harla. Pp 48- 72

$$P (E) = \frac{n (E)}{N (S)}$$

Donde: $N (S)$ = resultados igualmente probables del espacio muestra de un experimento. $n (E)$ = elementos de un hecho definido en este espacio muestra. $P (E)$ = probabilidad de este hecho.

Como este principio de razón insuficiente, depende del razonamiento lógico, generalmente no se tiene problema alguno si se tiene una moneda equilibrada o un dado no cargado, pero, en el mundo tal cual, no se cuenta con monedas perfectamente equilibradas por lo que es posible, basándose en este principio obtener resultados incorrectos al asignar probabilidades. Además, cuando se sabe que se tiene un dado o una moneda cargada, por ejemplo, si se utiliza la teoría clásica de la probabilidad, los resultados serán incorrectos.

2.1.3 La teoría de frecuencia relativa. En algunas ocasiones, un razonamiento a priori no funciona debido a la carencia de probabilidades iguales, en esas ocasiones, la teoría de la frecuencia relativa resulta de utilidad. Para esta teoría, la única forma de determinar probabilidades de hechos es por medio de experimentos repetitivos. Siguiendo el ejemplo clásico del lanzamiento de una moneda utilizado para ejemplificar la teoría clásica, la forma en que un partidario de la teoría de la frecuencia relativa actuaría sería lanzando una moneda un número grande de veces en las mismas condiciones, 100 por ejemplo, y si los resultados son 45 águilas y 55 soles, en consecuencia, la probabilidad para águila estaría expresada por la razón 45 / 100. Sin embargo, aunque la moneda estuviera equilibrada, podría suceder que no se obtuviera la probabi-

lidad verdadera, la probabilidad obtenida por medio de la utilización de la teoría de frecuencia relativa, se iría acercando a la probabilidad verdadera de $1 / 2$, a medida que el número de pruebas aumenta; la argumentación anterior conduce a la interpretación de probabilidad dentro de la teoría de la frecuencia relativa, en la que se considera que si un experimento se realiza en las mismas condiciones, n veces, y hay X resultados, dado X menor que n , en que ocurrió ese hecho, la probabilidad estará expresada por la razón:

$$\frac{X}{n}$$

Sin embargo, la verdadera probabilidad del hecho, estaría dada por la siguiente expresión:

$$P(E) = \lim_{n \rightarrow \infty} \frac{X}{n}$$

aquí, "lím" significa el valor que la función x/n adquirirá, cuando la condición $n \rightarrow \infty$ se cumpla.

Prácticamente, no sería posible obtener la probabilidad de un hecho con este límite, por lo que en realidad se obtiene una estimación aproximada de $P(E)$ dada una n grande. En otras palabras, esto quiere decir, que cuando se lanza una moneda o un dado equilibrados, es casi imposible que el resultado obtenido sea un águila, un sol y cada uno de los números del dado una vez; y si se lanza la moneda o el dado equilibrados un número grande de veces es de esperar, a largo plazo, o en promedio, que la mitad de las veces salga águila o sol y que cada número del dado aparezca $1 / 6$ parte de las veces; y es en este sentido que dentro de la teoría de la frecuencia relativa se dice que la probabilidad de obtener un número en un dado es de $1 / 6$.

Tanto la teoría clásica de la probabilidad como la teoría de la frecuencia relativa son consideradas como definiciones objetivas de probabilidad; la primera porque se basa en la deducción de un conjunto de supuestos y la segunda, porque la probabilidad de un hecho es determinada por observaciones experimentales repetidas. Sin embargo, no existe acuerdo en este último punto, ya que también se considera a la teoría clásica como subjetiva, pues la aceptación de los supuestos es un asunto meramente personal, no sujeto a comprobación experimental.

2.1.4 La Teoría Personalista.

La teoría clásica de probabilidad y la teoría de la frecuencia relativa, fácilmente resuelven problemas relacionados con el lanzamiento de una moneda, de un dado, o del resultado de una apuesta en la ruleta. No obstante, al enfrentar problemas de carácter único, sus procedimientos de resolución de los mismos no cuentan con la flexibilidad necesaria para dar soluciones a casos únicos, quedando un gran número de cuestiones fuera del alcance de los estadísticos seguidores de las teorías clásica y de frecuencia relativa. Asimismo, existe un gran número de casos en los que no es posible contar con toda la información necesaria para dar una solución utilizando la teoría clásica o la de frecuencia relativa y cuando la toma de decisiones se torna imperante, es la teoría personalista la que puede utilizarse.

El estadístico enmarcado en la teoría personalista o subjetiva considera a la probabilidad como un juicio personal en una proposición particular, asignando un peso entre cero y uno a un hecho específico según su grado de creencia en la posible ocurrencia del

mismo; también puede auxiliarse de la teoría clásica o de la de frecuencia relativa cuando no posea ninguna información que le permita asignar un peso específico; utilizando en ocasiones funciones conocidas que expresan el desarrollo de un fenómeno, en el caso de asignar probabilidades a priori.

Los teóricos subjetivos, admiten que las personas difieren al asignar probabilidades ante la misma evidencia, sin embargo, el subjetivista puede trabajar todos los problemas de los objetivistas y muchos más, llegando a las mismas conclusiones del objetivista cuando la cantidad de datos es lo suficientemente grande.

La estadística bayesiana, asunto fundamental de esta tesis, se encuentra clasificada dentro de la teoría personalista o subjetiva; y fue precisamente Baves el iniciador de dicho enfoque.

2.14.1 Bayes. Consideraciones preliminares.

Pierre Abailard y L.S. Berg (1981), consignan los siguientes datos y consideraciones sobre la vida y obra de Baves: nació en Londres, Inglaterra en 1702 y falleció en Tunbridge - Wells, Inglaterra, el 17 de abril de 1761. Perteneció a la primera generación de religiosos ingleses no conformistas. Fue educado privadamente y se cree que entre sus profesores de matemáticas se encontraba el famoso de Moivre. La Sociedad Real de Londres lo eligió como compañero en 1742. Sus obras fueron: "Divine benevolence" (1731) e "Introduction to the doctrine of fluxions" (1736), las cuales le fueron publicadas en vida. Su último trabajo fue un ataque a los fundamentos lógicos del cálculo newtoniano.

Su trabajo: "Essay Towards Solving a Problem in the - Doctrine of Chances" (1753) es el responsable de la relevancia contemporánea de Thomas Bayes; siendo así mismo, el primer intento de establecer una fundamentación de la inferencia estadística. Es considerado este ensayo, junto con "Ars Conjectandi de Jacques Bernoulli (1713) y "The doctrine of chances" de Abraham de Moivre (1713) como la base estructural de lo que ahora se conoce como - teoría de la probabilidad.

Su ensayo proporcionó los principios lógicos para inferir: dada la probabilidad de un evento, las probabilidades de eventos relacionados. En la obra de Bernoulli: "Ars Conjectandi", existía un conocimiento de como rechazar hipótesis estadísticas a la luz de nuevos datos, pero aún no se conocía la forma de establecer la probabilidad de una hipótesis estadística a la luz de los datos. La solución a este problema fue dada por Bayes cuando estableció que algunas veces la probabilidad de hipótesis estadísticas se encuentra dada aún antes que cualquier evidencia empírica sea observada. Posteriormente mostró cómo calcular la probabilidad de las hipótesis después de algunas observaciones. ,

La deducción del Teorema de Bayes, que se describe más adelante, podría ser considerada como una deducción hecha a partir de los axiomas de probabilidad básicos conocidos. Sin embargo, la más sorprendente característica del trabajo de Bayes es un argumento que no se deduce de ningún axioma estándar: en el caso de que no se sepa nada acerca de la probabilidad a priori de la hipótesis estadística que se está investigando, Bayes argumenta: " Yo tomaré por un hecho que la regla dada, relativa al evento M...es

también la regla a ser usada para cualquier evento por lo concerniente a la probabilidad de aquél del cual nada en absoluto es conocido anteriormente a cualquier ensayo realizado u observado con respecto a él. Y a tal evento le llamaré un evento desconocido" (24) Laplace aceptó el argumento de Bayes y gracias a esto desde 1774, la influencia de Bayes fue enorme, hasta que George Boole en "Laws of Thought" (1854) argumentó en contra; desde entonces, la teoría bayesiana es asunto de eterna controversia.

Leonard Savage en 1954, en "The Foundations of Statistics" rechaza la probabilidad objetiva e interpreta a la probabilidad como el grado de creencia personal de un individuo; aquí, la probabilidad a priori de una persona, es su creencia antes de que se realicen observaciones, y la probabilidad a posteriori, es aquella que tendrá la persona después de que las observaciones hayan sido hechas. Harold Jeffrey, británico, en 1961 afirmó en "Theory of Probability" que aún para la total ignorancia existe una distribución objetiva de grados de confianza para varias hipótesis. Savage y Jeffreys representan posiciones encontradas dentro de los estadísticos bayesianos contemporáneos. Por su parte, el mismo Thomas Bayes definió la probabilidad de cualquier evento como: "...la razón entre el valor al cual una expectativa que depende de la ocurrencia de un evento, debe ser computada, y el valor de la cosa esperada sobre su ocurrencia". (25) Y su definición puede

(24) BAYES, T., 1763. "Essay Towards Solving a Problem in the Doctrine of Chances". (Reprinted 1958). Biometrika. Vol. 45. P. 306.

(25) Bayes, T. Op cit. P. 299.

ser interpretada tanto desde el punto de vista subjetivo como del objetivo. A continuación, se revisarán los requisitos conceptuales sobre teoría de la probabilidad necesarios para la comprensión de los conceptos de probabilidad condicional y del teorema de Bayes.

2.2 Conceptos Básicos Sobre Probabilidad.

En primer lugar, se tiene que incertidumbre es el resultado de un proceso de cambio. Y si, un proceso de cambio puede - conducir a más de un resultado posible, se dice que los resultados son inciertos. Asimismo, si los resultados del experimento son inciertos, se dice que se trata de resultados aleatorios o estocásticos. La palabra experimento, puede referirse a uno físico, que puede realizarse un número X de veces; a pruebas efectuadas en condiciones similares o en el caso de un experimento aleatorio, como un conjunto de resultados en una secuencia de pruebas repetitivas. Se entiende por espacio muestra, a aquel conjunto universal, representado por la letra S , cuyos elementos representan todos los resultados posibles de un experimento.

En el caso de que se tengan dos resultados posibles en relación a la situación de un alumno en cuanto a su rendimiento escolar: aprobado o reprobado, se tendría:

$$S = \{A, R\}$$

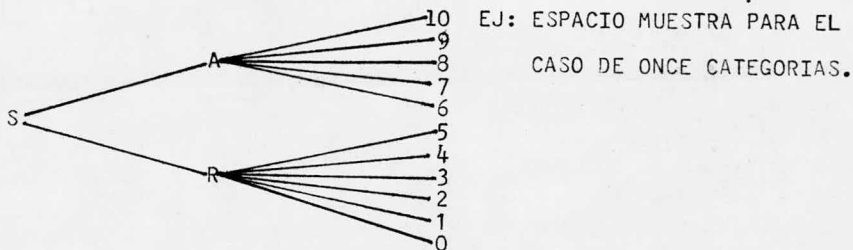
Si el caso es de 11 categorías de cero a 10 sería:

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Y en el de MB, B, S, NA sería: $S = \{MB, B, S, NA\}$

Un espacio muestra puede también representarse gráficamente, o por medio de un diagrama de árbol. A continuación, se representa por medio de un diagrama de árbol, un espacio muestra para el caso de

once categorías.



Siempre los resultados de S, cualquiera que sea la representación diagramática utilizada, deben ser mutuamente excluyentes y colectivamente exhaustivos. (Es mutuamente excluyente si no hay dos conjuntos que tengan puntos de muestra en común; es colectivamente exhaustivo si la unión de todos los resultados es igual al espacio muestra). Teniéndose así, que si dos o más resultados son mutuamente excluyentes no pueden ocurrir juntos, sino uno de ellos como máximo; y si un conjunto de hechos tiene las dos cualidades que se tratan, únicamente sucederá uno de los hechos. En consecuencia, teniendo el caso de 11 categorías para asignar al rendimiento de los alumnos, sólo puede ocurrir una de ellas en cada caso.

2.2.1 Primer axioma.

El rango en el que se encuentra la probabilidad va de 0 a 1 inclusive:

$$0 \leq P(A \setminus H) \leq 1 \quad (2.1)$$

donde se interpreta 0 como imposible y 1 como absolutamente verdadero bajo las condiciones establecidas. A representa cualquier

evento o alternativa, H representa las condiciones necesarias y el signo \leq significa menor o igual a...

2.2.2 Segundo axioma.

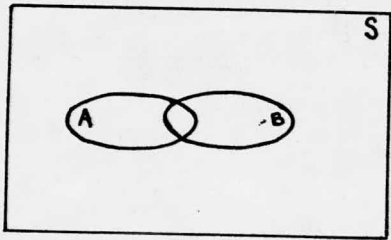
La probabilidad de todo espacio muestra es 1, o sea, $P(S) = 1$. Este axioma está basado en la convención existente en teoría de probabilidades por la cual se considera que si ha ocurrido un elemento del conjunto de hechos o más, la probabilidad asociada con S es 1.

2.2.3 Tercer axioma.

Si A y B son hechos disjuntos o sea: $A \cap B = \phi$, definidos en S , entonces $P(A \cup B) = P(A) + P(B)$. Este axioma es de suma importancia pues si de alguna forma puede asignarse un peso (probabilidad) a cada uno de los puntos de muestra de S , es posible obtener la probabilidad de cualquier hecho definido en S sumando los pesos de todos los puntos de muestra que son miembros del conjunto de hechos que se considera. Ec. (2.2)

2.2.4 Cuarto axioma.

Si A y B son dos eventos cualesquiera, definidos en S :



$$P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)} \quad (2.3)$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{de donde:} \quad (2.4)$$

$$P(A \cap B) = P(A) + P(B) - P(A \cup B). \quad (2.4a)$$

2.2.5 Probabilidad de un evento.

Como ya ha sido explicado anteriormente, un evento es un subconjunto del espacio muestra y el complemento de un evento A con respecto a S es el conjunto de todos los elementos de S que no están en A , lo cual se denota por: \bar{A} . Así, la probabilidad de dos eventos A y \bar{A} , complementarios, está dada por:

$$P(\bar{A}) = 1 - P(A). \quad (2.5)$$

2.3 Probabilidad Condicional.

Se llama probabilidad condicional, a aquélla relativa a que un evento B suceda, cuando se sabe que un evento A ocurrió. Se denota por:

$$P(B \setminus A)$$

que se lee: "La probabilidad de que B suceda, dado que A ocurrió." o "la probabilidad de B dado A ". Lo anterior puede ser clarificado mediante un ejemplo tradicional:

Considérese el evento X como obtener un cuadrado perfecto en el lanzamiento de un dado. Suponiendo que el dado esté contruido de tal manera que la obtención de cada cara sea igualmente probable. El espacio muestra S sería igual a $\{1, 2, 3, 4, 5, 6\}$; así, la probabilidad de que X suceda sería de $1/3$, pues los únicos cuadrados perfectos son 1 y 4. Ahora bien, ¿Cuál es la probabilidad de que en el lanzamiento de un dado se obtenga un cuadrado perfecto dado que se sabe que el resultado es un número mayor que 3? lo que también podría ser expresado como:

$$\text{¿Cuál es } P(X \text{ sea cuadrado perfecto} \setminus X > 3) ?$$

El espacio muestra con el que se está tratando en este nuevo problema es ahora $\{4, 5, 6\}$, el cual es un subconjunto de S . Así, podría plantearse la probabilidad condicional en la siguiente forma:

$$P(X \text{ sea CP} \mid X > 3) = \frac{P(4)}{P(4 \cup 5 \cup 6)} = \frac{1/6}{1/2} = 1/3$$

A continuación se establecerá el Teorema de Bayes a partir del siguiente razonamiento inductivo.

2.4 Teorema de Bayes.

Supóngase que S representa a la población de aspirantes a ingresar en una universidad, misma que se puede clasificar como sigue:

Tabla 3
Aspirantes a ingreso en una universidad.

	con beca (B)	sin beca (B)	Total
Hombres (M)	460	40	500
Mujeres (M)	140	260	400
Total	600	300	900

Al elegir al azar a uno de ellos, ¿qué probabilidad habrá de que sea hombre si se sabe que es becario?

Se trata de un caso de probabilidad condicional, que se podría plantear de la siguiente forma:

$$P(\bar{M} \mid B) = \frac{n(B \cap \bar{M})}{n(B)} \quad (2.6)$$

$$P(\bar{M} \mid B) = \frac{n(B \cap \bar{M}) / n(S)}{n(B) / n(S)} \quad (2.7)$$

$$P(\bar{M} \mid B) = \frac{P(B \cap \bar{M})}{P(B)} \quad (2.8)$$

¿Por qué se han establecido las igualdades anteriores?

Dado que el elegido es becario, es posible restringir la población al subconjunto de los becarios; por otro lado, ¿ a qué parte de la nueva población pertenece el elegido? Pues a aquellos que teniendo beca son hombres; esto es, a los que pertenecen a la intersección: becarios \cap hombres ($B \cap \bar{M}$). Por lo tanto, la probabilidad de que el elegido sea hombre, dado que es becario será:

$$P(\bar{M} \setminus B) = \frac{n(B \cap \bar{M})}{n(B)} = \frac{460}{600} = \frac{23}{30}$$

Por otra parte, si numerador y denominador de la anterior igualdad son divididos entre el número de elementos de la población total, la igualdad no se alterará, esto es:

$$P(\bar{M} \setminus B) = \frac{n(B \cap \bar{M}) / N(S)}{n(B) / N(S)}$$

Y siendo que, las nuevas expresiones para dividendo y divisor claramente representan probabilidades, se puede establecer que:

$$P(\bar{M} \setminus B) = \frac{P(B \cap \bar{M})}{P(B)}$$

Realizando los cálculos pertinentes:

$$P(B) = \frac{n(B)}{N(S)} = \frac{600}{900} = \frac{2}{3}$$

$$P(B \cap \bar{M}) = \frac{n(B \cap \bar{M})}{N(S)} = \frac{460}{900} = \frac{23}{45}$$

Se determina que como:

$$P(\bar{M} \setminus B) = \frac{P(B \cap \bar{M})}{P(B)} = \frac{23/45}{2/3} = \frac{23}{30}$$

que es el mismo resultado obtenido anteriormente. De la anterior igualdad se desprende que:

$$P(B \cap \bar{M}) = P(B) \cdot P(\bar{M} \setminus B) \quad (2.9)$$

Y análogamente puede demostrarse que:

$$P(B \cap \bar{M}) = P(\bar{M}) \cdot P(B \setminus \bar{M}) \quad (2.10)$$

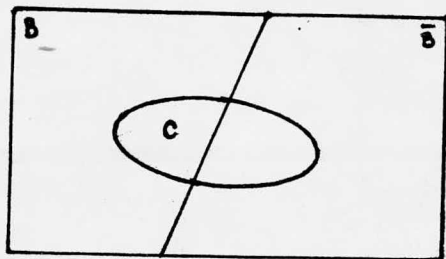
Estas dos últimas igualdades son de importancia fundamental en el Teorema de Bayes; pues en la primera de ellas se calcula la probabilidad de la intersección en función de $P(B)$ y de $P(\bar{M} \setminus B)$; en tanto que en la segunda, la probabilidad de la misma intersección se determina en función de $P(\bar{M})$ y de $P(B \setminus \bar{M})$.

En el ejemplo anterior, la probabilidad de seleccionar a alguien que fuese becario claramente era:

$$P(B) = \frac{600}{900} = \frac{2}{3}$$

Suponiendo que se sabe, además, que 36 de los becarios y 12 de los no becarios pertenecen a un círculo de estudios y sea C el evento de que el seleccionado sea miembro de dicho círculo: ¿qué probabilidad habrá de que el seleccionado sea becario si se sabe que pertenece al círculo?

Nuevamente se tiene el caso de una probabilidad condicional que puede ser imaginada más claramente mediante un diagrama de Venn Euler:



Lo que se pretende calcular ahora es la probabilidad de que el seleccionado esté en la intersección de los conjuntos B y C, esto es, en la zona sombreada del diagrama de Venn-Euler. Esto puede expresarse como sigue:

$$P(B \setminus C) = \frac{P(B \cap C)}{P(C)}$$

El evento C, esto es, que el seleccionado pertenezca al círculo será entonces:

$$C = (B \cap C) \cup (\bar{B} \cap C)$$

y por lo tanto:

$$P(C) = P(B \cap C) + P(\bar{B} \cap C)$$

De lo anterior:

$$P(B \setminus C) = \frac{P(B \cap C)}{P(B \cap C) + P(\bar{B} \cap C)}$$

y como

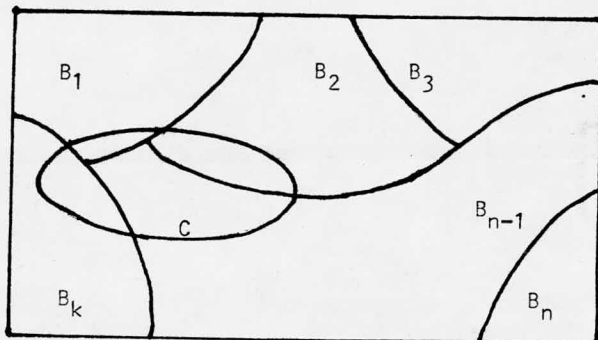
$$P (B \cap C) = \frac{36}{900} = \frac{3}{75}$$

$$P (\bar{B} \cap C) = \frac{12}{900} = \frac{1}{75}$$

$$P (B \setminus C) = \frac{3 / 75}{3 / 75 + 1 / 75} = \frac{3}{4}$$

Es evidente que el conocimiento de información adicional ha hecho que la probabilidad de que el seleccionado sea becario se incremente, lo que constituye uno de los fundamentos filosóficos del Teorema de Bayes, que en un sentido más amplio puede plantearse en la siguiente forma:

La información adicional, ya sea proveniente de consideraciones personales o del conocimiento de eventos relacionados - proporciona una ayuda no despreciable para el tratamiento de problemas inciertos; en el ejemplo anterior que se acaba de analizar $P (C)$ es de hecho la suma de probabilidades de todas las intersecciones posibles del evento C con los subconjuntos que forman la población; si la población se hubiera dividido en un número mayor de subconjuntos, por ejemplo, de acuerdo a la edad, sexo, promedio de calificaciones etc., se hubiesen presentado un mayor número de intersecciones, cada una de ellas de C con los subconjuntos de la población. Esto podría representarse gráficamente, de nuevo, mediante un diagrama de Venn-Euler.



Donde se ha denominado $B_1, B_2, \dots, B_k, \dots, B_n$ a los diferentes subconjuntos de la población. Entonces, la $P(C)$ estará dada por:

$$\begin{aligned}
 P(C) &= P(B_1 \cap C) + P(B_2 \cap C) + \dots + P(B_k \cap C) + \dots + \\
 &\quad + P(B_n \cap C) \\
 &= \sum_{i=1}^n P(B_i \cap C) \qquad (2.11)
 \end{aligned}$$

De donde, el teorema de Bayes para el caso de un evento de entre n posibles, dada una condición C se expresará:

$$P(B_k | C) = \frac{P(B_k \cap C)}{\sum_{i=1}^n P(B_i \cap C)} \qquad (2.12)$$

Y como:

$$P(B_i \cap C) = P(B_i) \cdot P(C | B_i) \qquad (2.13)$$

Entonces:

$$P (B_k \setminus C) = \frac{P (B_k \cap C)}{\sum_{i=1}^n P (B_i \cap C)} = \frac{P (B_k) P (C \setminus B_k)}{\sum_{i=1}^n P (B_i) \cdot P (C \setminus B_i)} \quad (2.14)$$

que constituye la expresión matemática del Teorema de Bayes.

Para ilustrar la aplicación de este teorema, se presenta el siguiente caso:

Tres miembros del profesorado de la Facultad de Psicología de cierta universidad, han sido nominados para la dirección de la misma. La probabilidad de que el Lic. Adame sea elegido, se estima en 0.3, la del Lic. Bravo en 0.5 y la del Lic. Coria en 0.2.

En el caso de que el Lic. Adame sea electo, se estima que los servicios administrativos tendrán una probabilidad de ser mejorados sustancialmente de 0.8, en el caso del Lic. Bravo de 0.1 y con el Lic. Coria de 0.4. Si después de cierto tiempo es notable la mejoría en los servicios administrativos, ¿cuál es la probabilidad de que el Lic. Coria haya resultado electo director de la facultad?

Si:

SM = evento de que los servicios han mejorado.

A = el Lic. Adame es director.

B = el Lic. Bravo es director.

C = el Lic. Coria es director

entonces:

$$P (C \setminus SM) = \frac{P (C \cap SM)}{P (A \cap SM) + P (B \cap SM) + P (C \cap SM)}$$

Como:

$$P (A \cap SM) = P (A) \cdot P (SM \setminus A) = 0.3 \times 0.8 = 0.24$$

$$P (B \cap SM) = P (B) \cdot P (SM \setminus B) = 0.5 \times 0.1 = 0.05$$

$$P (C \cap SM) = P (C) \cdot P (SM \setminus C) = 0.2 \times 0.4 = 0.08$$

de donde:

$$P (C \setminus SM) = \frac{0.08}{0.24 + 0.05 + 0.08} = 0.216$$

$$P (A \setminus SM) = \frac{0.24}{0.24 + 0.05 + 0.08} = 0.649$$

$$P (B \setminus SM) = \frac{0.05}{0.24 + 0.05 + 0.08} = 0.135$$

Como puede observarse, la probabilidad de que el Lic. Coria sea director, al igual que la del Lic. Bravo, si se toma en cuenta que los servicios mejoraron, es muy baja. En cambio para A dame, la situación es diferente.

Se concluye que, dado que los servicios mejoraron. La probabilidad de que el Lic. Adame haya sido electo, aumentó dramáticamente de 0.3 a 0.649. El peso de la evidencia adicional es notable.

2-5 ALGUNOS USOS DE LA ESTADISTICA BAYESIANA.

Actualmente los métodos bavesianos están siendo utilizados en prácticamente todos los campos: John Forester en 1973 (25) ingeniero industrial de la Universidad de Columbia elabora un manual sobre diversos usos de la estadística bavesiana en tomas de (25)FORESTER, J.,1973. "La Estadística en la Toma de Decisiones"

España: Bilbao.

decisiones empresariales. El doctor Gustavo Valencia, jefe de la materia de estadística bayesiana en la Facultad de Ciencias de la UNAM, ha realizado numerosas aplicaciones de los métodos bayesianos a problemas educativos y ecológicos, así como al desarrollo de nuevas vías en teoría del muestreo, paradigmas de investigación etc. Lawrence D. Phillips en 1974, en su obra "Bayesian Statistics for Social Scientists" propone procedimientos para pruebas de hipótesis a utilizar por diversos investigadores en ciencias sociales. En el Centro de Investigación Educativa de la Universidad de Leyden en Francia, el Dr. Hans F. Crombag se encuentra realizando investigación bayesiana en Evaluación Educativa junto con Henri Rouanet, Dominique Lépine y Jacqueline Pelnard-Considère. En los Estados Unidos existe un programa, auspiciado por Iowa Testing Program, que trata de encontrar nuevas aplicaciones de la metodología bayesiana a problemas de construcción y evaluación educativas.

Parece ser que la demostración lógico-matemática de Savage (26), de que el sistema bayesiano es el único sistema racional de toma de decisiones fue un punto medular que "puso de moda" a Thomas Bayes.

Actualmente, afirma Novick (27), los métodos bayesianos se encuentran disponibles para un número limitado de problemas, pues cuando se confrontan cuestiones más complejas, las matemáticas requeridas para producir soluciones se tornan muy complicadas

(26) SAVAGE, L.J. 1954, "The Foundations of Statistics". New York: Wiley. citado por: NOVICK, M.R. 1975 "Bayesian Methods in Educational Testing. A Third Survey". Advances in Psychological and Educational Measurement. London: Ed. by D.N.M. de Grujiter. Pp.17-32.

(27) NOVICK, M.R. 1975. Op cit. Pp. 33.

no obstante, a partir de la última década, afirma Novick, el incremento de la utilización de la metodología bayesiana ha sido - exponencial.

Actualmente, ¿dónde se encuentra la confrontación método clásico vs. método bayesiano? El método clásico es bien entendido y conocido, así como sus debilidades, lo que posibilita una cuidadosa atención en sus aplicaciones. Aún más, el método clásico tiene un amplio registro de aplicaciones adecuadas. Por su parte, el método bayesiano proporciona una teoría más rica, susceptible de ser aplicada a un mayor número de problemas. Asimismo, los métodos bayesianos carecen de una validación empírica amplia, sin embargo, ésta se incrementa cada día.

I I I M E T O D O P R O P U E S T O .

3.1 ANTECEDENTES.

En educación, en evaluación educativa y particularmente en evaluación referida a un criterio, asunto de este trabajo, Bayes ha encontrado un lugar preponderante. Como afirma Jackson en 1975 (28), en los campos social y educativo, en donde no es posible contar con una evidencia empírica abrumadora para la toma de decisiones, es indispensable utilizar todos los datos objetivos disponibles así como "nuestro conocimiento" y "nuestra experiencia"; por lo tanto y dadas las características del modelo bayesiano, parece ser que es el más adecuado. Con referencia a cuál es el camino a seguir cuando la evidencia no es abrumadora, señala Jackson cuatro rutas a seguir derivadas de situaciones diferentes: 1) Cuando el investigador desea mejorar sus propias decisiones. Es necesario que el tomador de decisiones formule sus expectativas a priori como una distribución de probabilidades cuidadosa y consistentemente; esto puede realizarse utilizando un programa por computadora con procedimientos iterativos que forzan al usuario a ser consistente pues lo enfrentan repetidamente con varias consecuencias de

(28) JACKSON, P.H. 1975. "The Philosophy and Methodology of Bayesian Inference". Advances in Psychological and Educational Measurement. London: Ed. by Gruijter. Pp.3-16

sus respuestas, por ejemplo conjuntos de intervalos de confianza a priori. 2) Cuando los datos serán utilizados para persuadir o convencer a otra persona, cuyas expectativas a priori son diferentes de las del investigador. Esta situación se presenta cuando los experimentos son planeados para desafiar una teoría existente; en este caso las creencias previas del experimentador no son las importantes sino la de la persona o personas que "juzgarán el caso"; para lo cual el experimentador tendrá que dar mayor "evidencia favorable" que si hubiera utilizado su propia distribución a priori. 3) Cuando el usuario de la información es desconocido en el momento en que los datos son analizados. En algunas ocasiones es éste el problema del reporte científico; un procedimiento razonable es reportar la función más probable, la cual resume la evidencia de la muestra, junto con una indicación de la distribución a posteriori, la cual resultaría de la adopción de una variedad de distribuciones a priori diferentes. 4) Finalmente, cuando las inferencias deben ser realizadas para un número determinado de personas interesadas, cada una de ellas con sus opiniones a priori y donde no es posible acumular evidencia suficiente para llegar a un consenso. Esta situación sucede cuando algún asunto se somete a un análisis público o cuando un test de amplitud nacional es aplicado. Las creencias del examinador acerca de la habilidad del examinado pueden ser bastante diferentes de las que el examinado tenga de sí mismo por no mencionar las de sus maestros, padres y abuelos etc. En un caso como éste se podría optar por tomar un punto de vista neutral en la distribución a priori. Una distribución de este tipo también podría ser utili-

zada cuando la cantidad de la información a priori parece muy endeble, tanto que el esfuerzo de cuantificar no sería justificable. En este contexto se habla de una información a priori vaga o no informativa. Al respecto Jeffreys (29) afirma que "nuestra ignorancia" debe expresarse como una distribución uniforme con parámetros de rango $(-\infty, \infty)$.

Rouanet, Dominique y Pelnard-Considère (30), investigadores franceses presentan un método bayesiano simplificado, que sin ayuda de computadoras permite tratar los datos educativos y encontrar una distribución a posteriori de distribuciones a priori de diversos tipos, tales como "ambiente favorecedor", pruebas de "inteligencia", y métodos de enseñanza.

Hambleton y Novick (31) utilizan metodología bayesiana en instrucción modular monitoreada; el paradigma en este tipo de instrucción es: el estudiante domina el módulo o no lo domina. La decisión basada en el método bayesiano consiste en permitir que el alumno avance al siguiente módulo o retenerlo. ¿Cuándo se deberá permitir que el alumno avance de un módulo al siguiente? En los casos extremos, cuando el nivel de dominio sea o muy alto o muy bajo, aparentemente no habría ningún problema; sin embargo, en los casos dudosos existe un beneficio muy grande en u-

(29)JEFFREYS, M 1961. "Theory of Probability" Oxford: Clarendon Press (third edition).

(30)ROUANET, H. et al. "Bayes-Fiducial procedures as practical substitutes for misplaced significance testing? an application to educational data". Advances in Psychological and Ed. M. op.cit.

(31)NOVICK, M.R 1975. Op. Cit.

tilizar la información previa tanto de las ejecuciones del estudiante en cuestión como la de los registros de los estudiantes que han pasado por la misma situación, así como la información colateral que proporcione el sistema, por ejemplo: el conocimiento de que las funciones de pérdida dependen altamente de la relación entre los módulos.

Como es sabido, uno de los principales responsables del desenvolvimiento de la evaluación referida a un criterio en los Estados Unidos, fue el proyecto Talent, que proponía el desarrollo e implementación de diversas opciones para mejorar la calidad de la instrucción. Una de las características de los nuevos programas fue que estaban redactados en términos de objetivos de instrucción, siendo la meta máxima de un programa por objetivos el proporcionar un sistema educativo susceptible de ser adaptado a las necesidades del aprendiz, por lo que, en consecuencia, surgieron numerosos sistemas de enseñanza individualizada.

Los tests referidos a un criterio han sido definidos de múltiples maneras, sin embargo, una definición aceptable sería la proporcionada por Glaser y Nitko en 1971 (32): "Un test referido a un criterio es aquél que se encuentra deliberadamente construido de tal manera que sus mediciones sean directamente interpretables en términos de estándares de ejecución especificados.

(32)GLASER, R and NITKO. "Measurement in learning and instruction" En R.L Thorndike (Ed.), Educational measurement. Washington:American Council on Education. 1971. P. 653.

Los estándares de ejecución son generalmente especificados definiendo algunos dominios de tareas que el estudiante debe realizar. Muestras representativas de las tareas a dominar son organizadas en un test. Las mediciones son tomadas y son utilizadas para realizar afirmaciones acerca de la ejecución relacionada con el dominio de cada individuo".

Hambleton, Swaminathan y Algina (33), proponen asimismo la utilización de la metodología bayesiana a una amplia gama de problemas de la evaluación referida a un criterio.

Actualmente, en la Universidad de Iowa, el Dr. Melvin Novick y colaboradores, se encuentran desarrollando nuevos usos de la metodología bayesiana a la evaluación educativa y problemas derivados de la evaluación referida a un criterio; asimismo, en la Dirección de Evaluación de la Secretaría de Educación Pública, al mando del Ing. Estelio Baltazar Cadena, se empiezan a conocer los métodos bayesianos y se planean diversos estudios de aplicación.

3.2 PROCEDIMIENTO.

Se propone un procedimiento estadístico basado en el Teorema de Bayes, para la predicción de calificaciones o notas en marcadas en un tipo de evaluación referida a un criterio.

El método propuesto pretende, dada una calificación inicial, predecir una calificación final, lo cual podría ser de utilidad para: tomar decisiones de diversa índole que permitirían

(33)HAMBLETON, R.K. et al. 1975. "Some contributions to the theory and practice of criterion-referenced testing" en Advances in Psychological and Educational Measurement. London.

la aplicación de cursos de nivelación y planear acciones correctivas a mediano y largo plazo para una administración educativa más eficiente.

La población de este estudio corresponde a las calificaciones (N=332) de los grupos de Fisicoquímica de un maestro de la Escuela Superior de Ingeniería Química e Industrias Extractivas del Instituto Politécnico Nacional, que abarcan el período del 11 de noviembre de 1978 al 13 de febrero de 1982.

El marco de referencia para la asignación de calificaciones fue el de relación a la perfección, relacionado con el grado de dominio demostrado por los sujetos en relación con los objetivos de la materia, redactados conductualmente de acuerdo con los lineamientos enunciados en el capítulo I.

Las calificaciones fueron asignadas inicialmente en forma de porcentajes, sin embargo, posteriormente fueron transformadas a letras de acuerdo a la siguiente regla:

CUADRO I EQUIVALENCIA DE CALIFICACIONES.

0 - 59 %	N.A.
60 - 79 %	S
80 - 94 %	B
95 - 100 %	M.B.

Dicha transformación se llevó a cabo con objeto de hacer el método directamente aplicable al sistema universitario de asignación de calificaciones.

Las expectativas a priori podrían haber sido formuladas

entre otras posibilidades, de acuerdo con las del maestro, con - las del mismo alumno, o de acuerdo con una a priori no informativa; sin embargo, se optó por tomar un punto de vista más o menos neutral en la distribución a priori y se reportó la función más probable como la proporcionada por la evidencia de la muestra.

A continuación, se enunciará en forma de pregunta y específicamente, la meta fundamental de este estudio, así como los diferentes pasos que se siguieron en el desarrollo del mismo en forma sucinta; posteriormente, se desarrollarán los mismos secuencialmente en forma más amplia.

¿Cuál es la probabilidad de que un estudiante obtenga en la materia de Fisicoquímica, ya sea MB, B, S o NA como calificación definitiva, si la primera calificación parcial obtenida por él, es MB, B, S o NA?

El procedimiento puede resumirse como sigue:

I.-Elección de una muestra aleatoria.

II.-Determinar para la muestra, las frecuencias de cada una de las calificaciones iniciales C_I y de las definitivas, C_D (Tabla IV).

III.-Determinar para la muestra con base en las calificaciones definitivas C_D , las correspondientes probabilidades de obtener cada una de ellas, $P(C_{Di})$. Estos valores constituirán las probabilidades a priori en el Teorema de Bayes.

IV.-Determinar las probabilidades de intersección

$P(C_I \cap C_D)$ según:

$$P(C_I \cap C_D) = P(C_I \setminus C_D) P(C_D)$$

mismas que servirán de base para el cálculo de las probabilidades a posteriori:

$P(C_D \setminus C_I)$ de acuerdo al Teorema de Bayes según:

$$P(C_D \setminus C_I) = P(C_I \cap C_D) / P(C_I)_M$$

donde $P(C_I)_M$ es la probabilidad marginal de C_I . Lo anterior implica la aplicación del Teorema de Bayes, ya que:

$$P(C_{Dk} \setminus C_I) = \frac{P(C_{Dk}) P(C_I \setminus C_{Dk})}{\sum_{i=1}^n P(C_{Di}) P(C_I \setminus C_{Di})} \quad (3.1)$$

$$P(C_{Dk} \setminus C_I) = \frac{P(C_{Dk} \cap C_I)}{\sum_{i=1}^n P(C_{Di} \cap C_I)} \quad (3.1 a)$$

$$P(C_{Dk} \setminus C_I) = \frac{P(C_{Dk} \cap C_I)}{P(C_I)_M} \quad (3.1 b)$$

Es entonces claro que la probabilidad marginal de C_I está dada por:

$$P(C_I)_M = \sum_{i=1}^n P(C_{Di} \cap C_I) \quad (3.1 c)$$

esto es, es la suma de probabilidades de las intersecciones de las diferentes C_{Di} con una calificación C_I dada. (Ver tablas V y VI).

V.-Asignando un número a cada posible calificación:

NA	S	B	MB
1	2	3	4

y de acuerdo con los principios de "valor esperado" o "expectativa matemática" expresados por Ya Lun Chou (34): V.1 Determinar, a partir de las probabilidades a priori, el valor esperado de la calificación definitiva:

$$C_D^* = (1) P(NA_D) + (2) P(S_D) + (3) P(B_D) + (4) P(MB_D) \quad (3.2)$$

V.2 Determinar la calificación definitiva esperada, para cada calificación del primer examen parcial:

$$C_D^{**} = (1) P(NA_D \setminus C_I) + (2) P(S_D \setminus C_I) + (3) P(B_D \setminus C_I) + (4) P(MB_D \setminus C_I) \quad (3.3)$$

VI.-Establecer en cuántos casos de la población concuerdan los valores de C_D , con los valores esperados C_D^* y C_D^{**} , para cada calificación C_I (Ver tabla VII).

VII.-Responder a la pregunta: ¿es el Teorema de Bayes un mejor método de estimación de la calificación definitiva que el estimador de probabilidad clásica tradicional? Esta pregunta se resolverá mediante un análisis gráfico y una prueba χ^2 .

I.-Determinación de la muestra. La muestra fue determinada de acuerdo con las siguientes consideraciones:

Con base en el teorema del límite central que establece que: "Cuando se dan n variables aleatorias independientes $X_1, X_2, X_3, \dots, X_n$ que tienen todas ellas la misma distribución (Chou, Y, L, 1977. "Análisis Estadístico". México: Interamericana Págs. 118-120.

quiera que esta sea) entonces:

$$X = X_1 + X_2 + X_3 \dots + X_n$$

es normal y asintótica. La media μ y la varianza σ^2 de X son:

$$\mu = \mu_1 + \mu_2 + \dots + \mu_n = n \mu_i$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2 = n \sigma_i^2$$

donde μ_i y σ_i^2 son la media y la varianza de X_i ."

"Cuando se dice "misma distribución", se quiere decir que:

1.- Tanto X_1 como X_2 tienen distribuciones similares; es decir, ambas tienen distribuciones normales o rectangulares o binomiales, etc.

2.- Las medias μ_1 y μ_2 son iguales, $\mu_1 = \mu_2$

3.- Las varianzas σ_1^2 y σ_2^2 son iguales, $\sigma_1^2 = \sigma_2^2$

O que: supóngase una población de N estudiantes. Selecciónese (con reemplazo) un estudiante y digamos que X_1 sea su variable aleatoria de CI. Selecciónese un segundo estudiante cuya variable aleatoria de CI sea X_2 ; entonces, si X_1 y X_2 provienen de la misma población X_1 y X_2 tendrán la misma distribución. Si se seleccionan 3 estudiantes (con reemplazo) se puede interpretar que tienen tres va-

riables aleatorias independientes X_1, X_2, X_3 , que tienen la misma distribución."

"En general, cuando se toma una muestra aleatoria de tamaño n de una población, se tendrán n variables aleatorias independientes

$$X_1, X_2, \dots, X_n$$

que tendrán todas la misma distribución..." (35)

Tomando en cuenta el Teorema del Límite Central, principio fundamental en que se basa la teoría del muestreo, se procedió a seleccionar el método más apropiado de elección de la muestra dentro de una amplia gama de los mismos. Así, se eligió el de muestreo para proporciones, en virtud de su sencillez y de que los parámetros necesarios son fácilmente obtenidos de una "premuestra" de la población.

El procedimiento consiste en clasificar los eventos individuales en dos clases, C y C' por ejemplo, donde C representa al subconjunto de los alumnos aprobados y C' a su complemento. Se ha convenido en un margen de error d , en la proporción estimada P , de unidades pertenecientes a la clase C y un pequeño riesgo α de que el error real sea mayor que d , parámetros que se fijaron en los siguientes valores:

$$d = 0.10$$

$$\alpha = 0.05$$

(35)YAMANE, T., 1974. "Estadística". México:Harla, S.A. de C.V. Págs. 89 y 90.

Una muestra de 20 casos, reveló una proporción de alumnos aprobados (p) de 75 %, de donde la de no aprobados (q) será de 25 %.

De lo anterior, en la fórmula para determinación del tamaño de la muestra para proporciones (36):

$$n = \frac{\frac{t^2 pq}{d^2}}{1 + \frac{1}{N} \left(\frac{t^2 pq}{d^2} - 1 \right)} \quad (3.4)$$

donde:

p = proporción de casos aprobados en la muestra.

q = proporción de casos reprobados en la muestra.

t = valor de desvío normal (desviación estándar) correspondiente al nivel (α) de confianza deseado.

d = desviación permitida como aceptable en la predicción de la calificación.

Si: p = 0.75, q = 0.25, t = 1.96 (α = 0.05), d = 0.10 y como N = 332 :

$$n = \frac{\frac{1.96^2 \times 0.75 \times 0.25}{0.1^2}}{1 + \frac{1}{332} \left(\frac{1.96^2 \times 0.75 \times 0.25}{0.1^2} \right)} = 59.18$$

(36) COCHRAN, W.G. 1971. "Técnicas de Muestreo". México: CECSA Págs. 109 y 110.

por lo que la muestra se fijó en 60 casos.

II.-Determinación de frecuencias de calificaciones para la muestra. Los resultados de una muestra de 60 casos, elegida al azar, de una población de 332, se resumen en la tabla 4:

Tabla 4
Elementos de la muestra según C_I v C_D

$C_I \backslash C_D$	NA	S	B	MB	Total
NA	10	7	1	0	18
S	5	5	4	0	14
B	2	6	14	6	28
MB	0	0	0	0	0

III.-Determinación de las probabilidades a priori. Se calcularán a partir de las frecuencias de calificaciones definitivas C_D que aparecen en la muestra, sus probabilidades correspondientes, $P(C_{Di})$.

De la tabla anterior, se desprende que:

$$P(N A_D) = \frac{n(N A_D)}{n(S)} = \frac{18}{60} = 0.300$$

$$P(S_D) = \frac{n(S_D)}{n(S)} = \frac{14}{60} = 0.233$$

$$P(B_D) = \frac{n(B_D)}{n(S)} = \frac{28}{60} = 0.467$$

$$P(MB) = \frac{n(MB_D)}{n(S)} = \frac{0}{60} = 0$$

IV.- Determinación de las probabilidades de intersección:

$$P(C_I \cap C_D)$$

Como:

$$P(C_I \cap C_D) = P(C_I \setminus C_D) P(C_D) \quad \vee$$

$$P(C_I \setminus C_D) = \frac{n(C_I \cap C_D)}{n(C_D)}$$

y según los datos de la tabla 4 y del punto III, se tiene que:

$$P(NA_I \cap NA_D) = \frac{10}{18} \times 0.3 = 0.1667$$

y así sucesivamente, estos resultados pueden resumirse en la tabla 5.

Tabla 5

Probabilidades de intersección de las calificaciones de la muestra.

$C_D \backslash C_I$	NA_I	S_I	B_I^-	MB_I
NA_D	0.1667	0.1167	0.0167	0
S_D	0.0832	0.0832	0.0666	0
B_D	0.0334	0.1001	0.2335	0.1001
MB_D	0	0	0	0
$P(C_I)_M$	0.2833	0.300	0.3168	0.1001

Donde $P(C_I)_M$ es la probabilidad marginal de cada calificación inicial:

$$P(C_I)_M = \sum P(C_I \cap C_{D_i})$$

Ahora, se pueden calcular las probabilidades a posteriori, según el teorema de Bayes, para cada condición de C_I , usando para ello las probabilidades marginales recién obtenidas, ya que:

$$P(C_{D_i} \setminus C_I) = P(C_{D_i} \cap C_I) / P(C_I)_M \quad \dots (3.5)$$

obsérvese sin embargo que dado que:

$$P(C_{D_i} \cap C_I) = P(C_{D_i}) P(C_I \setminus C_{D_i}) \quad \dots (3.6)$$

entonces (3.5) puede tomar la forma:

$$P(C_{D_i} \setminus C_I) = P(C_{D_i}) \frac{P(C_I \setminus C_{D_i})}{P(C_I)_M} \quad \dots (3.7)$$

que es la forma clásica de la expresión matemática del teorema de Bayes, en donde:

$P(C_{D_i})$ expresa el estado de conocimiento antes de que se conozca C_I , y por lo mismo se denomina probabilidad a priori.

$P(C_{D_i} \setminus C_I)$ establece el estado de conocimiento sobre C_D después de aprender acerca de C_I , por lo que se le conoce como probabilidad a posteriori.

Las ecuaciones (3.5) y (3.7) son entonces equivalentes y pueden usarse indistintamente.

La generalización del procedimiento señalado en las ecuaciones (3.5), (3.6), y (3.7), conduce a una forma alternativa

de la expresión del Teorema General de Bayes plasmada en la ecuación (3.1):

$$P(C_{Dk} \setminus C_I) = \frac{P(C_{Dk}) P(C_I \setminus C_{Dk})}{\sum_{i=1}^n P(C_{Di}) P(C_I \setminus C_{Di})}$$

De acuerdo con (3.5) y según la tabla V:

$$P(NA_D \setminus NA_I) = \frac{0.1667}{0.2833} = 0.5884$$

y así subsecuentemente, los resultados se resumen en la siguiente tabla:

Tabla 6
Probabilidades a posteriori

\tilde{E}	$P(C_D \setminus NA_I)$	$P(C_D \setminus S_I)$	$P(C_D \setminus B_I)$	$P(C_D \setminus MB_I)$
NA_D	0.5884	0.3890	0.0527	0
S	0.2937	0.2773	0.2102	0
B_D	0.1179	0.3337	0.7371	1
MB_D	0	0	0	0

V.-Cálculo de Valores Esperados Tradicional y Bavesiano. A continuación se determinan con las probabilidades de calificaciones definitivas, $P(C_{D_i})$ el valor esperado de probabilidad tradicional y con los valores de calificaciones definitivas dadas ciertas calificaciones iniciales, $P(C_{D_i} \setminus C_I)$, los valores esperados Bavesianos, asignando un número a cada calificación.

NA	S	B	MB
1	2	3	4

V.1 Para el estimador tradicional.

$$C_D^* = (1)XP(NA_D) + (2)XP(S_D) + (3)XP(B_D) + (4)XP(MB_D)$$

Utilizando los resultados obtenidos en el punto III:

$$\begin{aligned} C_D^* &= (1) \times (0.300) + (2) \times (0.233) + (3) \times (0.467) + (4) \times (0) \\ &= 0.300 + 0.466 + 1.401 = 2.167 \sim S \end{aligned}$$

V.2 Para el estimador Bayesiano.

$$C_D^{**} = (1)P(NA_D \setminus C_I) + (2)P(S_D \setminus C_I) + (3)P(B_D \setminus C_I) + (4)P(MB_D \setminus C_I)$$

para cada calificación inicial:

$$C_{D_{NA_I}}^{**} = (1) \times (0.5884) + (2) \times (0.2937) + (3) \times (0.1179) + (4) \times (0) = 1.529 \sim S$$

$$C_{D_{S_I}}^{**} = (1) \times (0.3890) + (2)(0.2733) + (3) \times (0.3337) + (4) \times (0) = 1.9367 \sim S$$

$$C_{D_{B_I}}^{**} = (1)(0.0527) + (2)(0.2102) + (3)(0.7371) + (4)(0) = 2.684 \sim B$$

$$C_{D_{MB_I}}^{**} = (1)(0) + (2)(0) + (3)(1.000) + (4)(0) \approx 3.00 \sim B$$

VI.-Calificaciones iniciales y definitivas para la población.

Las calificaciones de toda la población, se distribuyen de acuerdo a las siguientes particiones:

Tabla 7

Elementos de la población según C_I v C_D

$C_I \backslash C_D$	NA_D	S_D	B_D	MB_D	Total
NA_I	36	38	12	0	86
S_I	42	51	22	0	115
B_I	9	21	57	1	88
MB_I	1	1	39	2	43
Total	88	111	130	3	332

El estimador tradicional C_D^* , predice que, sea cual sea la calificación original, la calificación definitiva que es de esperarse será S_D , esto es cierto en el caso de que la calificación inicial sea S, dudoso para $C_I=NA$ y no se cumple cuando C_I es B o MB.

El estimador Bayesiano, C_D^{**} , en contraste, establece que cuando C_I es NA, se esperará como definitiva NA o S, con muy ligera preferencia sobre ésta última; que cuando C_I es S, la calificación esperada al final será S; que cuando la C_I es B, resultaría B como definitiva, y por último, que cuando C_I es MB, la calificación definitiva que habrá de esperarse será B; todo esto está de acuerdo con lo observado en la población. Esto significa, entonces, que el estimador Bayesiano es aparentemente, más certero - que los estimadores tradicionales.

VII.- Ahora bien, ¿es el Teorema de Bayes un mejor método de estimación de la calificación definitiva que el estimador de probabilidad clásica tradicional?

VII.1 Dado que las calificaciones usadas, NA, S, B y MB o numéricamente hablando 1, 2, 3 y 4 son intrínsecamente discretas, los estimadores, ya sea el tradicional o los bayesianos deberán tener valores congruentes con aquéllos. Esto es:

$$C_D^* = 2.167 \longrightarrow 2.0$$

$$C_{NA_I}^{**} = 1.529 \longrightarrow 2.0$$

$$C_{D_{S_I}}^{**} = 1.937 \longrightarrow 2.0$$

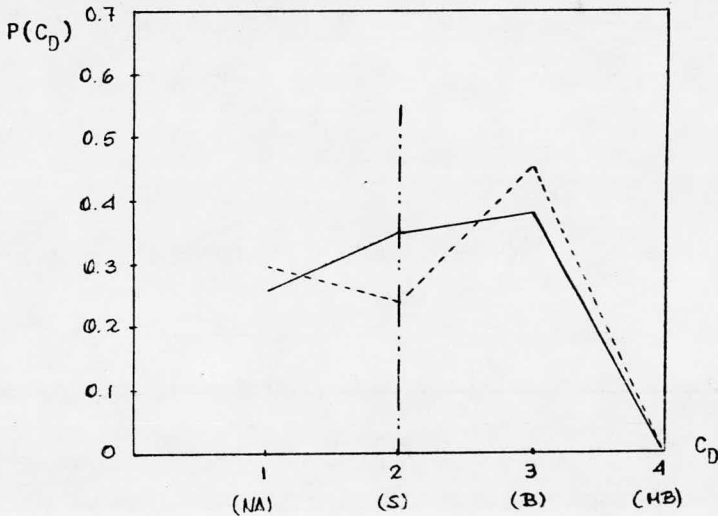
$$C_{D_{B_I}}^{**} = 2.684 \longrightarrow 3.0$$

$$C_{D_{MB_I}}^{**} = 3.000 \longrightarrow 3.0$$

Todo lo anterior se ilustra en las figuras 3.1 y 3.2. En la figura 3.1, se indican las probabilidades directas de cada calificación definitiva, tanto de la población (línea continua) como de la muestra (línea punteada), así como el valor esperado (línea vertical).

Figura 3.1

Probabilidad de calificación definitiva vs. calificación definitiva.



De esta figura es nuevamente evidente que la calificación definitiva esperada, obtenida por el método tradicional, es 2.0 ó sea S. Si se observa su posición en la figura 3.1, se ve que la línea vertical no coincide con la calificación más probable de la población, y por lo mismo el poder predictivo de este estimador es bastante relativo.

Por otro lado, en las cuatro gráficas de la figura 3.2 se observa, confirmando lo que ya se había determinado previamente, que los estimadores bayesianos, ó sea, las calificaciones esperadas obtenidas utilizando el teorema de Bayes, coinciden rigurosamente con las calificaciones más probables de la población. En consecuencia, el poder predictivo de los estimadores bayesianos es mucho mayor, en términos absolutos, que el del estimador tradicional.

Sin embargo, con objeto de reforzar aún más esta conclusión, se procederá a realizar una prueba de hipótesis χ^2 (chi cuadrada).

En la página siguiente se presentará la figura 3.2 con las siguientes gráficas:

Gráfica A. Probabilidad vs. Calificación Definitiva para Calificación Inicial NA.

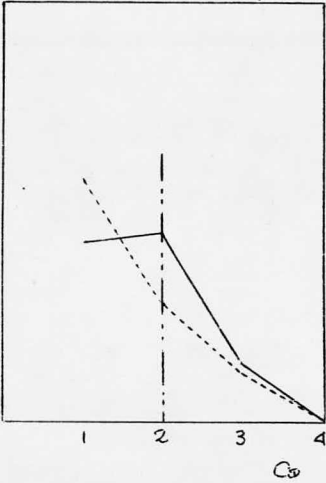
Gráfica B. Probabilidad vs. Calificación Definitiva para Calificación Inicial S.

Gráfica C. Probabilidad vs. Calificación Definitiva para Calificación Inicial B.

Gráfica D. Probabilidad vs. Calificación Definitiva para Calificación Inicial MB.

Probabilidad vs. Calificación definitiva para calificación inicial NA.

$C_I = NA$

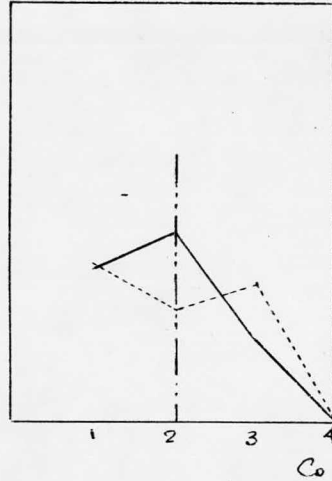


$P(C_D)$

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

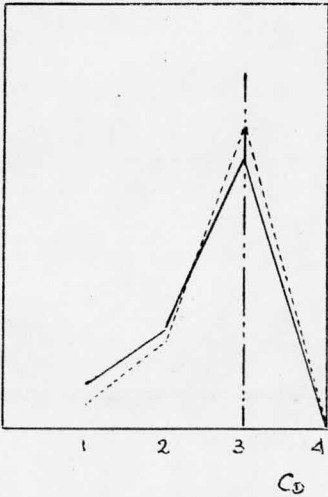
Probabilidad vs. Calificación definitiva para calificación inicial S.

$C_I = S$



Probabilidad vs. Calificación definitiva para calificación inicial B.

$C_I = B$

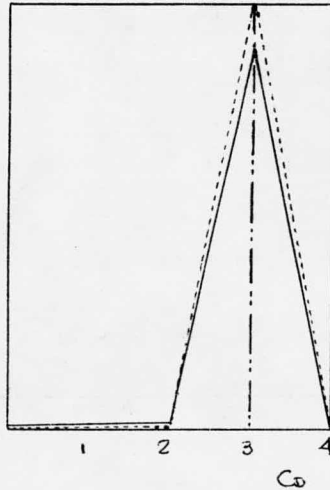


$P(C_D)$

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

Probabilidad vs. Calificación definitiva para calificación inicial MB.

$C_I = MB$



C_D

VII.2 Prueba de Significancia. Con objeto de confirmar la evidencia mostrada en el análisis gráfico de los resultados, se decidió utilizar la prueba de hipótesis conocida con el nombre de χ^2 (Chi cuadrada), que es una medida de la diferencia entre las frecuencias observadas y teóricas. Cuando estas coinciden, $\chi^2 = 0$. Esta prueba sirve para determinar cuándo las discrepancias entre las frecuencias observadas y las esperadas son mayores que lo que puede esperarse por mero azar.

La distribución χ^2 se define como:

$$\chi^2 = \sum_{i=1}^c \frac{(f_i - F_i)^2}{F_i} \quad (3.9)$$

donde f_i = frecuencia observada en la clase "i"

F_i = frecuencia esperada en la misma clase.

c = número de clases.

Cuando se trata de r muestras de c clases, la distribución χ^2 toma la forma:

$$\chi^2 = \sum_1^r \sum_1^c \frac{(f_i - F_i)^2}{F_i} \quad (3.9 a)$$

Un concepto relacionado con esta prueba es el de Grados de Libertad, que puede considerarse como el número de valores de frecuencia, por cada muestra, que pueden adoptar, libremente, diversos valores; evidentemente, cuando se trata de una sola muestra,

En la tabla anterior:

$$f_{I_{C_D}} = n (C_I)_{\text{Mue.}} \cdot P(C_{D_i} \setminus C_I)_{\text{Pob.}}$$

Lo que conduce a la siguiente tabla:

Tabla 9

Frecuencias observadas v esperadas. (Para dos categorías).

categoría	NA _D		S _D		B _D		MB _D	
	f _i	F _i	f _i	F _i	f _i	F _i	f _i	F _i
clásico	16	18	20	14	23	28	1	0
bayesiano	17	17	15	18	28	19	0	6

De donde:

$$\chi^2 = \sum_1^r \sum_1^c \frac{(f_i - F_i)^2}{F_i} = \frac{(16 - 18)^2}{18} + \frac{(20 - 14)^2}{20} + \frac{(23 - 28)^2}{28}$$

$$\frac{(1 - 0)^2}{1} + \frac{(17 - 17)^2}{17} + \frac{(15 - 18)^2}{18} + \frac{(28 - 19)^2}{19} + \frac{(0 - 6)^2}{6} = 9.9$$

Para este caso:

$$\text{Grados de Libertad} = (r - 1)(C - 1) = 1 \times 3 = 3$$

De las tablas de χ^2 , con 3 grados de libertad:

$$P(\chi^2 > 9.9) = 0.02$$

Por lo que la hipótesis nula debe rechazarse, de donde es entonces evidente que ambos métodos de predicción no son igualmente efectivos, lo que desde luego coincide con lo señalado en las páginas anteriores.

Ahora bien, la anterior prueba señala que ambos sistemas de predicción: el clásico y el bayesiano, no son equivalentes, pero no indica cual de ellos es más acertado. Sin embargo, en el análisis gráfico que aparece en la página 89, se observa que el estimador bayesiano es el que se ajusta más adecuadamente a los datos de la población.

3.2 Conclusiones.

El procedimiento estadístico basado en el Teorema de Bayes que ha sido presentado en este estudio tiene como finalidad fundamental el predecir la calificación final a partir de una calificación inicial.

En este caso particular, el conocimiento utilizado como base de los cálculos fue el proporcionado por la evidencia de una muestra de 60 casos de una población de 332 sujetos, alumnos de diferentes cursos de Fisicoquímica de un mismo profesor de la ESIQIE del IPN. De hecho, se trabajó con una distribución experimental que señala la tendencia en obtención de calificaciones de una materia, en una escuela, en un tipo de evaluación referida a un criterio. Del mismo modo, pudieron haber sido elegidas como fuentes de información: el registro de calificaciones de cada uno de los alumnos, o las opiniones que sobre él presentaran profesores o el mismo estudiante.

Tanto el método de probabilidad clásica como el bayesiano, se basaron en los datos proporcionados por la muestra. Asimismo, los resultados proporcionados por ambos procedimientos fueron comparados con los datos surgidos de la población. En consecuencia, resulta impresionante la exactitud de la inferencia bayesiana, misma que se torna evidente en el análisis gráfico explicitado en la figura 3.2 de la página 88.

Utilizando el método tradicional, se infiere para el caso en cuestión, que sea cual sea la calificación original, la calificación definitiva será S. Aquí existe una inexactitud, ya que esto es únicamente verdadero cuando la calificación inicial es S, posible cuando la calificación inicial es NA y no se cumple cuan-

do la calificación inicial es NA y no se cumple cuando es B o MB. Con el método bayesiano, la predicción es más exacta como aparece explícitamente en la página 86.

El manejo estadístico bayesiano de las calificaciones, proporcionará únicamente resultados útiles cuando los procesos - que subyacen a la evaluación educativa han sido realizados cuidadosamente.

La exactitud de la predicción, utilizando el método de inferencia bayesiana o cualquier otro, dependerá de que se cuente con un patrón común de aplicación e interpretación. Si una calificación de MB ó 10 significa que el alumno se esforzó mucho a juicio del maestro, no obstante, que su ejecución sólo cubrió el 60 % de los objetivos, por ejemplo, La predicción realizada con esta información no resultará exacta cuando en mediciones y evaluaciones posteriores los criterios cambien.

Siguiendo en la misma línea del estudio presentado, como sugerencia para otros posibles usos del sistema bayesiano propuesto, está el del problema específico de determinar si un alumno entraría o no a un curso de postgrado u obtendría una beca para la cual existen muchos aspirantes y pocas plazas. El procedimiento que generalmente se sigue es solicitar la mayor cantidad posible de información por parte del solicitante, aplicar un solo examen de confiabilidad dudosa y tomar una decisión "a sentimiento". En este caso, la utilización del Teorema de Bayes con el método - propuesto para predicción de calificaciones sería de gran utilidad, pues proporcionaría un sistema confiable de inferencia para la toma de decisiones.

El procedimiento bayesiano propuesto está dado para el caso de un curso completo de un semestre, pero también podría ser utilizado con una unidad de programa, o con un módulo. La probabilidad a posteriori obtenida podría ayudar a determinar si el alumno puede pasar al siguiente módulo o no.

La inferencia de calificaciones realizada con el sistema bayesiano propuesto parece permitir con base en una predicción más exacta que las decisiones de administración educativa mejoren.

Probabilísticamente hablando, parece ser, basándose en los resultados del estudio que al enfrentar problemas de carácter único, los procedimientos de solución de la teoría de probabilidad clásica no tienen la flexibilidad indispensable para dar solución efectiva a casos particulares.

Los usos y finalidades de la enseñanza y particularmente de la medición y evaluación educativas han dependido a lo largo de la historia de la humanidad, de la cosmovisión de los seres que habitamos este planeta, de los valores contextuales de la sociedad, de los propósitos particulares de las personas que realizamos las mediciones y las evaluaciones y desafortunadamente de la improvisación. El sistema bayesiano permite tomar decisiones a pesar de la incertidumbre, haciendo a un lado la improvisación, incorporando toda la información disponible y considerada en un momento dado pertinente por el tomador de decisiones. Debido a lo anterior, se considera provechosa la realización de nuevas experiencias de aplicación del sistema bayesiano que pongan de manifiesto nuevas ventajas y limitaciones de la metodología bayesiana, ratificando o no las conclusiones preliminares de estudios como el presente.

Nomenclatura.

Mayúsculas:

- (A B): condicionalidad; evento A, dado B.
 C : número de clases.
 C : calificación.
 F : frecuencia esperada.
 G.L.: grados de libertad.
 N : número total de elementos en la población.
 P : probabilidad.
 S : espacio muestral.

Minúsculas:

- d : desviación permitida como aceptable en la predicción de la calificación.
 f : frecuencia observada.
 n : número total de elementos en la muestra.
 p : proporción de casos aprobados en la muestra.
 r : número de muestras.
 t : valor de desviación estándar correspondiente al nivel () de confianza deseado.

Subíndice:

- D : definitiva.
 I : inicial.
 i : iésimo valor.

k : kaésimo valor; valor específico de un índice variable.

M : marginal.

Supraíndice:

* : valor esperado de probabilidad clásica.

** : valor esperado de probabilidad bayesiana.

— : testa, evento complementario.

Griegas:

α : nivel de confianza deseado.

μ : media de la población.

Σ : operador sumatoria.

σ^2 : varianza de la población.

χ^2 : chi cuadrada.

A 1 G L O S A R I O

Administración educativa: es aquél proceso que consta de las siguientes etapas:

- 1.-Identificación del problema basándose en las necesidades.
- 2.-Determinación de los requisitos y alternativas de solución.
- 3.-Selección, entre opciones, de las estrategias de solución.
- 4.-Implantación de las opciones escogidas.
- 5.-Determinación de la eficiencia de la ejecución.
- 6.-Revisión continua.

Aleatorios (resultados).: si los resultados de un experimento son inciertos, esto es que pueden ser más de uno, se dice que son aleatorios.

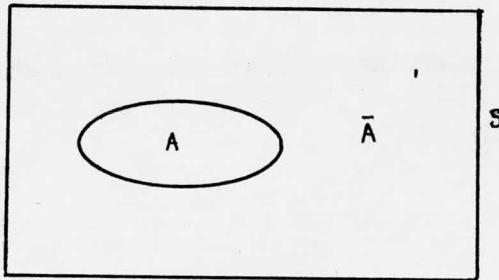
Asintótica (distribución): dícese de la curva que se acerca indefinidamente a una recta, sin llegar a encontrarla.

Axioma: postulado que se acepta sin ser probado.

Colectivamente exhaustivo (resultados): cuando la unión de todos los resultados es igual al espacio muestra.

Complemento (Conj.):el complemento de un evento A con respecto a S, es el conjunto de todos los elementos de S que no están en A.

Se denota por los símbolos A' o \bar{A} .



Confiabilidad: "Es la consistencia que muestran los datos a través del tiempo (dadas las mismas condiciones), el grado de acuerdo entre observadores independientes o la consistencia de los datos a través de diferentes investigaciones", (Castro, 1976).

Conjunto: es una colección de cosas particulares.

Conjunto vacío o espacio nulo: es aquel subconjunto del espacio muestra que no tiene elementos. Se denota por el símbolo ϕ .

Curva normal de probabilidad: es una gráfica idealizada de ciertas distribuciones de probabilidad cuya validez crece cuando N es muy grande y que sigue la expresión:

$$p(x) = \frac{1}{\sqrt{2\pi} \sigma^2} e^{-x^2 / 2\sigma^2}$$

Desviación estándar: medida de variabilidad o dispersión de la puntuación en un grupo determinado. La desviación estándar es la raíz cuadrada del promedio de las desviaciones cuadradas respecto

de la media aritmética del grupo. La desviación estándar proporciona la unidad en términos de la cual se expresan las puntuaciones estándar.

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n}}$$

Diseño de la prueba: formulada la hipótesis, el próximo punto es el que trata de los pasos destinados a la comprobación o no de las hipótesis presentadas. Este paso es el diseño de la prueba. Comprobación quiere decir el grado de probabilidad estimado en que las observaciones estén de acuerdo con las hipótesis propuestas.

Distribución: es el arreglo relativo de un conjunto de números; un conjunto de valores de una variable y las frecuencias de cada valor.

Distribución de frecuencias: ordenamiento de un conjunto de puntuaciones correspondientes a un grupo de individuos en el que los valores de las puntuaciones posibles se disponen en orden de superior a inferior y se indica el número de personas que obtiene cada puntuación.

Error de medición: es la cantidad en que cualquier medición específica difiere de la puntuación 'verdadera', hipotética del individuo en lo tocante a la cualidad que se está midiendo. Puesto que ningún procedimiento de medición es perfectamente exacto, cada uno tiene un componente de error.

Espacio muestra: es aquél conjunto universal representado por la letra S, cuyos elementos representan todos los resultados posibles de un experimento. Es la reunión de todos los puntos muestra en un conjunto.

Estocásticos (resultados): si los resultados de un proceso de cambio son inciertos, o sea que hay más de un resultado posible se dice que son estocásticos.

Evaluación: proceso dentro del cual quedan comprendidos la interpretación de puntajes y la toma de decisiones para determinar si son o no adecuados para ciertos fines.

Evaluación diagnóstica: es aquélla que tiene como propósito principal la búsqueda de información acerca de la naturaleza de algún tipo de problema o de alguna insuficiencia en un programa, persona, proceso, componente, etc.

Evaluación formativa: es aquella evaluación que tiene como propósito el suministro de la información que conduce a la modificación y continuo mejoramiento de la unidad o ente que se está evaluando.

Evaluación sumativa: es aquélla que tiene como propósito extraer una decisión clara que puede ser binaria o múltiple o de calificación acerca de un programa, persona o componente, en función - de su mérito o valor, generalmente en relación con algún hito importante.

Evento simple: es aquel evento de un conjunto que contiene sólo un elemento del espacio muestra.

Evento compuesto: es aquél que puede expresarse como la unión de eventos simples.

Experimento: puede referirse a uno físico que puede realizarse un número X de veces; a pruebas efectuadas en condiciones similares o en el caso de un experimento aleatorio, como un conjunto de resultados en una secuencia de pruebas repetitivas.

Función: es una asociación de exactamente un objeto de un conjunto (el rango) con cada objeto de otro conjunto (el dominio). Ej.: el área de un círculo es función del radio. La expresión $y = 3x^2 + 7$ define a y como función de x cuando se especifica que el dominio es, por ejemplo, el conjunto de los números reales; y es entonces una función de x , un valor de y es asociado con cada valor de número real de x multiplicando el cuadrado de x por 3 y agregando 7 (el rango de esta función es el conjunto de todos los números reales no menores que 7), y se dice que x es la variable independiente o argumento de la función (y se llama variable dependiente o valor de la función). Para denotar una función se utilizan símbo

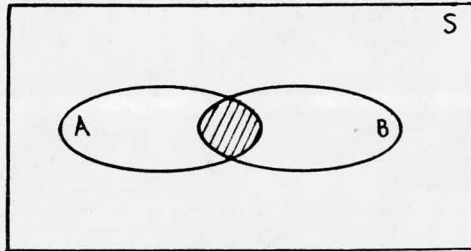
los como: f , F , etc. La función de valores correspondientes a x se denota por $f(x)$, $F(x)$ etc. y se lee "f de x" ó "f función de x".

Grados de libertad: el número de valores de frecuencia por cada muestra que pueden adoptar, libremente, diversos valores.

Para una muestra: G.L. = $C - 1$; donde C = número de categorías o clases.

Para r muestras: G.L. = $(r - 1) (C - 1)$

Intersección: la intersección de dos eventos A y B , es el evento que contiene todos los elementos en común de A y B . Se denota por el símbolo $A \cap B$.



Iterativos (procedimientos): proceso en el cual se repite la acción encaminada al logro de los fines, tomando en cuenta los resultados anteriores y acercándose a la solución del problema por medio de aproximaciones sucesivas.

Halo o aura, error de.: el término designa la tendencia de algunos evaluadores a fundar las estimaciones de una persona que está siendo evaluada en una opinión general favorable respecto de ese individuo y a no distinguir el grado de posesión de rasgos específicos.

Hechos disjuntos: si $A \cap B = \emptyset$. Cuando dos eventos no pueden ocurrir al mismo tiempo.

Hecho o evento: es un subconjunto del espacio muestra.

Simple: un evento que consiste de un punto en el espacio.

Compuesto: aquél que está formado por dos o más puntos.

Hipótesis: es una afirmación que indica una relación específica entre dos o más variables. Es una respuesta tentativa que se propone a una pregunta o problema planteado.

Hipótesis alternativa: es aquella que establece que la variable independiente tiene efecto sobre la variable dependiente.

Hipótesis nula: es aquella que establece que la variable independiente no tiene efecto sobre la variable dependiente y que por lo tanto no se espera que existan diferencias en los valores de la variable dependiente en las distintas condiciones experimentales.

Medición: es la asignación de números a cosas de acuerdo a reglas. Es el proceso que involucra: la especificación de objetivos de aprendizaje, la construcción de instrumentos de medición, la administración y la asignación de puntajes a las pruebas.

Modelo: es una representación que demuestra las relaciones entre variables pertinentes usadas para entender mejor y controlar a las situaciones reales.

Objetivo: "Es un propósito expresado en un enunciado que describe el cambio propuesto en el alumno cuando haya terminado exitosamente una experiencia de aprendizaje" (Mager, 1969)

Permutación: es un arreglo de todo o parte de un conjunto de objetos. El número de permutaciones de n distintos objetos es $n!$. El número de permutaciones de n distintos objetos tomados r a la vez es:

$${}_n P_r = \frac{n!}{(n-r)!}$$

Probabilidad: es la medida y manera de expresar la incertidumbre y los riesgos asociados a ella.

Bayesiana: es el grado de creencia en la posible ocurrencia de un hecho.

A priori: es la creencia de un individuo sobre la ocurrencia de un hecho antes de que se realicen observaciones.

A posteriori: es la creencia de un individuo sobre la ocurrencia de un hecho después de que se realizan observaciones.

Condiciona: es aquella relativa a que un evento B suceda cuando se sabe que un evento A ya ocurrió. Se denota por: $P(B \setminus A)$.

Marginal: es la suma de probabilidades de las intersecciones de los diferentes valores de la variable dependiente con cada uno de los valores de la variable independiente. Su fórmula general sería la siguiente; tomando a X como variable independiente y a Y como variable dependiente.

$$P(X_k) = \sum_{i=1}^n P(Y_i \cap X_k)$$

Punto muestra: cada resultado de un experimento.

Puntuación bruta: es una puntuación expresada en las unidades en las que se le obtuvo originalmente, es decir, en los puntos obtenidos en la aplicación de la prueba.

Razón insuficiente (principio): en un experimento al azar, los resultados son igualmente probables.

Resultado aleatorio o estocástico: si los resultados del experimento son inciertos.

Resultados inciertos: si un proceso de cambio puede conducir a más de un resultado posible se dice que los resultados son inciertos.

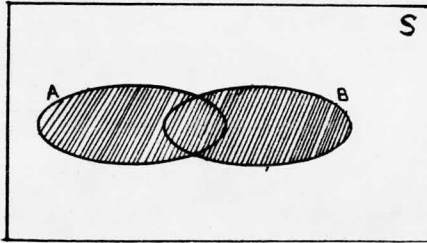
Realimentación: es aquella función de un sistema que compara el producto o resultado obtenido con los criterios de rendimiento establecidos. El objetivo básico es el de controlar el rendimiento del sistema, suministrando formaciones básicas que permiten los progresivos ajustes a fin de eliminar las diferencias entre los criterios y el producto.

Sistema: es un ente o conjunto de entes compuesto de partes que están organizadas en un todo destinado al logro de uno o más propósitos generales o específicos.

Subconjunto:dícese del conjunto formado por uno o algunos de los elementos de un conjunto.

Teorema:conclusión general propuesta para ser probada sobre la base de ciertas hipótesis previas.

Unión:la unión de dos eventos A y B, es el evento que contiene todos los elementos que pertenecen a A o B o ambos y se denota por $A \cup B$



Si A y B son eventos o afirmaciones, cuando A y B no son mutuamente excluyentes, esto es, pueden ocurrir al mismo tiempo, la probabilidad de A o B está dada por:

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

Cuando A y B son mutuamente excluyentes:

$$(A \cap B) = 0 \quad \text{y:}$$

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

Validez:es la eficiencia de una prueba para representar, describir o pronosticar el atributo que le interesa al aplicador.

Valor esperado o expectativa matemática: medida de tendencia central utilizada para caracterizar distribuciones de probabilidades. "El valor esperado de una variable aleatoria X es un promedio ponderado de los valores que puede asumir X con probabilidades para los valores de X como pesos." (Chou, 1977).

Variable: es un símbolo utilizado para representar un miembro no especificado de algún conjunto. Cualquier miembro del conjunto es un valor de la variable y el conjunto en sí mismo es el rango de la variable. Si el conjunto tiene sólo un elemento, la variable es una constante.

Variable aleatoria: es una función X cuyo rango es un conjunto de números reales $(-\infty, +\infty)$, cuyo dominio es el espacio muestral (conjunto de resultados) S de un experimento, y para el cual el conjunto de todos los s en S , para el cual $X(s) = x$, es un evento si x es un número real. También puede ser definida como la función P de probabilidad, sobre subconjuntos adecuados de un conjunto T ; los puntos de T son eventos elementales y cada conjunto en el dominio de P un evento.

Variable aleatoria continua: es aquella variable aleatoria cuyo espacio muestral es discreto, esto es contiene un número finito o infinito numerable de puntos muestrales que forman una continuidad.

Variable aleatoria discreta: es aquella cuyo espacio muestral es discreto, esto es, por ejemplo, el espacio muestral cuyos puntos se pueden colocar en una correspondencia uno a uno con los

números naturales. $(1 - \infty)$.

Referencias bibliográficas.

ABAD, S.A. y L.A. SERVIN. 1981. "Introducción al muestreo". México: Limusa.

ABAILARD, P y L.S. BERG. 1981. "Dictionary of scientific biography". New York: Charles Scribner's sons. Págs. 530-532.

ANDERSON, R.C. y G.W. FAUST. 1977. "Psicología Educativa". México: Trillas.

ASHBAGH, E.J. 1924. 1924. "Reducing the variability in teachers' marks". Journal of Educational Research, 9: Págs. 185-198.

BAYES, T. 1763. "Essay towards solving a problem in the doctrine of chances". (Reimpreso 1958) Biometrika No 1. 45.

BLOOM, B.S., et al. 1975. "Evaluación del Aprendizaje". Volumen I. Argentina: Troquel.

CASTRO, L. 1976. "Diseño experimental sin estadística". México: Trillas.

COCHRAN, W.G. 1976. "Técnicas de muestreo". México: CECSA.

CHOU, Y.L. 1977. "Análisis estadístico". México: Interamericana.

- DADOURIAN, H.M. 1925. "Are examinations worth the price". EUA: School and Society.
- DU BOIS, P.H.A. 1970. "History of psychological testing". Boston: Allyn and Bacon Inc.
- EBEL, R.K. 1965. "Measuring educational achievement". Englewood Cliffs, N.J.:Prentice-Hall, Inc. Citado por STANLEY y HOPKINS op cit.
- EDWARDS, A.L. 1968. "Experimental design in psychological research". Third Edition. EUA : Holt, Rinehart and Winston, Inc.
- EVES, H. 1964. "An introduction to the history of mathematics". Third Edition. EUA.:Holt, Rinehart and Winston, Inc.
- FALLS, J.D. 1928. "Research in secondary education". Kentucky School Journal, 6: Págs 42-46. Citado por Dadourian, op cit.
- FORESTER, J. 1973. "La estadística en la toma de decisiones". España:Deusto.
- FREEMAN, H. 1970. "Introducción a la inferencia estadística". México:Trillas.
- GLASS, G.V. y J.C. STANLEY. 1970."Statistical methods in edu-

tion and psychology". Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

GLASSER, W. 1969. "Schools without failure". New York: Harper and Row Publishers.

GLASSER, R. y A.J. NITKO. 1971. "Measurement in learning and instruction". En THORNDIKE, R.L. (Ed.), "Educational measurement" (2da. Ed.) Washington: American Council on Education.

GUTRIE, W.K. 1955. "Los filósofos griegos". México: Bréviarios Fondo de Cultura Económica.

GOODWIN, W.M. y W.K. LE BOLD. 1975. "A discriminant analysis approach to placement". En Advances in Psychological and Educational Measurement. London: Ed. D.N.M. de Gruijter.

GUTTENTAG, M. 1977. "Evaluation and society". Personality and Social Psychology Bulletin. Vol 3.

HAMBLETON, R.K. et al. 1975. "Some contributions to the theory and practice of criterion referenced testing". En Advances in Psychological and Educational Measurement". London: Ed. D.N. de Gruijter.

HILGARD, E.R. y D.C. MARQUIS. 1972. "Condicionamiento y apren-

dizaje". México: Trillas.

JACKSON, P.H. 1975. "The philosophy and methodology of bayesian inference". En *Advances in Psychological and Educational Measurement*. Op cit.

JEFFREYS, S.M. 1961. "Theory of probability". Oxford: Clarendon Press. Third edition.

JOHNSON, F.W. 1911. "A study of high school grades". *EUA.: School Review* 19.

JUECES, 12: 5-6, Versión del Rev Juan.

LEWIS, CH. 1975. "Marginal distributions for the estimation of proportions in m groups". *Psychometrika* Vol 28.

MAGER, R. 1969. "La confección de objetivos para la enseñanza". Cuba: Ministerio de Educación.

MAGER, R. 1975. "Medición del intento educativo". Argentina: Guadalupe.

MANN, H. 1845. "Report of the annual examining committee of the Boston grammar and writing schools". *Common School Journal*, 7: Págs. 326-336. Citado por DADOURIAN, H.M. op cit.

MEYER, M. 1908. "The grading of students". *Science*, 27: Págs 243-250. Citado por STANLEY, J.C. v K.D. HOPKINS. 1976. op cit.

MOOD, A.M. v F.A. GRAYBILL. 1972. "Introducción a la teoría de la estadística". Madrid: Aguilar.

MORICE, E. 1975. "Diccionario de estadística". México: CECSA.

MOYNIHAM, P. 1971. "Seek parity of educational achievement". Moynihan urges". *EUA : Report on Educational Research*. (march) 3 (5).

NOVICK, M.R. 1975. "Bayesian methods in educational testing: A third survey". En *Advances in Psychological and Educational - Measurement*". op cit.

NOVICK, M.R. 1970. "Bayesian considerations in educational - information systems". *EUA: Educational Testing Service: Págs. 77-89.*

OBREGON, S.I. 1977. "Teoría de la probabilidad". México: Limusa.

PETERS, F.R. 1956. "Measurement of informal educational achievement". *School Review*, 64: Págs 227-232.

- PHILLIPS, L.D. 1974. "Bayesian statistics for social scientists".
EUA, New York: Thomas v Crowell Company.
- PLUTCHIK, R. 1967. "Foundations of experimental research". EUA:
Harpers Experimental Psychology Series under the Editorship of
H.Philip Zeigler.
- RICE, J.M. 1897. "The futility of the spelling grind" :I Forum,
23: Págs 163-172. Citado por STANLEY J.C. Y.K.D. HOPKINS.op
cit.
- ROMERO, L. 1976. Apuntes de clase de la materia psicometría. -
México: UAM.
- ROUANET, H. et al. "Bayes fiducial procedures as practical subs
titutes for misplaced significance testing". "An application to
educational data". Advances in Psychological and Educational -
Measurement. London: Gruijter.
- SAVAGE, L.J. 1954. "The foundations of statistics". New York:
Wiley. Citado por NOVICK, M.R. 1975,op cit.
- STANLEY, J.C. v K.D. HOPKINS. 1976. "Educational and psycholog
ical measurement and evaluation". Englewood Cliffs, N.J.:Prentice-Hall.
- SMITH, E.R. v R.W. TYLER. 1942. "Appraising and recording stu-

dent progress". New York: Harper and Row, Publishers. Citado por Stanley y Hopkins, op cit.

STARCH, D. 1913. "Reliability and distribution of grades". *EUA: Science*, 38.

STONE, C.W.. 1908. "Arithmetical abilities and some factors - determining them". New York: Columbia University Press. Citado por Stanley y Hopkins, op cit.

SUND, R.D. y A.J. PICARD. 1976. "Objetivos conductuales y medidas de evaluación". México: Trillas.

THORNDIKE, E.L. 1904. "Introduction to the theory of mental and social measurements". New York: Teachers College, Columbia University. Citado por DU BOIS, P.H. 1970. Boston: Allyn and Bacon Inc.

THORNDIKE, R.L. (Ed) 1971. "Educational measurement". (2da.Ed) Washington: American Council on Education.

THORNDIKE, R.L. y E. HAGEN. 1970. "Tests y técnicas de medición en psicología y educación". México: Trillas.

WHITE, E.E. 1886. "The elements of pedagogy". New York: American Book Company. Citado por DU BOIS, op cit.

WHITE, H.B. 1932. "Testing as an aid to learning". *EUA: Educational Administration and Supervision*, 18.

WILLEMSEN, E. 1974. "Understanding statistical reasoning". "How to evaluate research literature in the behavioral sciences". *EUA: San Francisco: W.H. Freeman and Company*.

YAMANE, T. 1974. "Estadística". México: Harla. S.A. de C.V.

YOUNG, R.K. y D.J. VELDMAN. 1968. "Introducción a la estadística aplicada a las ciencias de la conducta". México: Trillas.