



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

DISEÑO DE UN INSTRUMENTO PARA EVALUAR CURSOS DE EDUCACION CONTINUA EN COMPUTO.

T E S I S
QUE PARA OBTENER EL TITULO DE
A C T U A R I O
P R E S E N T A
LUIS ANTONIO HERAS MONROY



DIRECTOR DE TESIS: ACT. HORTENSIA CANO GRANADOS



2002



Universidad Nacional
Autónoma de México



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.



DRA. MARÍA DE LOURDES ESTEVA PERALTA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

Diseño de un instrumento para evaluar cursos de educación continua en cómputo.
realizado por Luis Antonio Heras Monroy
con número de cuenta 9332946-6, quien cubrió los créditos de la carrera de: Actuarial

Dicho trabajo cuenta con nuestro voto aprobatorio.

A t e n t a m e n t e

Director de Tesis
Propietario

Act. Hortensia Cano Granados

Propietario

Act. Leticia Daniel Orana

Propietario

Act. Maria Guadalupe Izquierdo Dyrzo

Suplente

Act. Benigna Cuevas Pinzón

Suplente

Act. Maria Aurora Valdes Michell

Consejo Departamental de Matemáticas

M. en C. Jose Antonio Flores Diaz

UNIVERSIDAD DE LA PAZ

Gracias Padre Mio

por demostrarme que **no** estoy solo
y por hacer de mis **sueños una realidad...**

Gracias Papá, Mamá, **Hermano** y Abuelita
sé que nada ha sido **fácil**
pero al final, todo el **esfuerzo** ha valido la pena...

Gracias amigos míos
a los que están y a los que **ya no** están

...

Gracias a ti
que por el cariño incondicional que me has brindado
llevas ya un lugar muy **especial** en mi corazón...

Gracias a los sinodales y al **personal** de la DGSCA
que han formado parte **en la** realización
de este trabajo...

Y muy en particular

agradezco infinitamente a la **Maestra** y Act. Hortensia Cano Granados
ya que con su **conocimiento, apoyo** y **paciencia** he logrado culminar esta tarea.

**DISEÑO DE UN INSTRUMENTO PARA EVALUAR CURSOS DE
EDUCACIÓN CONTINUA EN CÓMPUTO**

ÍNDICE

INTRODUCCIÓN	i
---------------------------	----------

CAPÍTULO I

LA EVALUACIÓN Y LA MEDICIÓN

1.1 La evaluación.....	1
1.2 La medición.....	2
1.3 Clasificaciones de la evaluación.....	3
1.4 Técnicas de evaluación.....	5
1.5 Funciones de evaluación.....	10

CAPÍTULO II

LA EDUCACIÓN CONTINUA Y LA DGSCA

2.1 La Educación Continua y sus objetivos.....	12
2.1.1 Clasificación de los actos académicos de la Educación Continua.....	13
2.1.2 Tipos de actos académicos de la Educación Continua.....	14
2.2 La Dirección General de Servicios de Cómputo Académico (DGSCA).....	17
2.2.1 Algo de historia.....	17
2.2.2 Funciones de la DGSCA.....	18
2.2.2.1 Docencia en Cómputo.....	20
2.2.2.2 Apoyo a la Investigación.....	20
2.2.2.3 Desarrollo en sistemas computacionales.....	20
2.2.2.4 Telecomunicaciones.....	21
2.2.3 La Educación Continua en la DGSCA.....	22
2.2.4 Los Centros de Cómputo.....	23
2.2.5 Otros servicios de la DGSCA.....	24

CAPÍTULO III

INSTRUMENTOS DE MEDICIÓN

3.1 Definición de Instrumento de Medición	25
3.2 Algunas consideraciones	26
3.3 Características de un instrumento de medición	27
3.3.1 La Confiabilidad	27
3.3.1.1 Factores que influyen en la confiabilidad	28
3.3.1.1.1 El calificador	28
3.3.1.1.2 El contenido del instrumento	29
3.3.1.1.3 El tiempo	29
3.3.1.2 Evaluación de la Confiabilidad	29
3.3.1.2.1 Método de Test-Retest	30
3.3.1.2.2 Método de Formas Equivalentes	31
3.3.1.2.3 Método de Mitades Partidas	32
3.3.1.2.4 Coeficiente α de Cronbach	33
3.3.1.2.5 Coeficiente KR20	36
3.3.2 La Validez	37
3.3.2.1 Evaluación de la Validez	38
3.3.2.1.1 Validez de contenido	38
3.3.2.1.2 Validez ligada a criterio	39
3.3.2.1.2.1 Validez concurrente	39
3.3.2.1.2.2 Validez predictiva	40
3.3.2.1.3 Validez de constructo	42
3.4 Otros aspectos por considerar	44

CAPÍTULO IV

DISEÑO Y CONSTRUCCIÓN DE UN INSTRUMENTO DE MEDICIÓN

4.1 Procedimiento para construir un instrumento de medición	45
4.2 Objetivo del instrumento.....	46
4.2.1 Beneficios de una evaluación	46
4.3 ¿Qué instrumento de medición emplear?	47
4.3.1 La encuesta	47
4.3.2 La observación	48
4.3.3 El Cuestionario	51
4.3.3.1 Tipos de cuestionarios.....	51
4.3.3.1.1 Cuestionario estructurado directo	53
4.3.3.1.2 Cuestionario no estructurado directo	53
4.3.3.1.3 Cuestionario estructurado indirecto.....	53
4.3.3.1.4 Cuestionario no estructurado indirecto.....	54
4.3.3.2 Composición de un cuestionario.....	54
4.4 Definición de las variables a evaluar.....	55
4.4.1 Tipos de variables.....	55
4.4.1.1 Aspectos por considerar.....	55
4.4.1.1.1 Indicadores de una buena enseñanza	57
4.4.1.1.2 Procedimientos de evaluación	60
4.4.1.1.3 Agentes evaluadores	60
4.4.1.2 Establecimiento y alcance de las variables a evaluar.....	61
4.5 Definición de los reactivos y de la escala de medición	62
4.5.1 Escala nominal.....	63
4.5.2 Escala ordinal	63

4.5.3 Escala de intervalo	64
4.5.4 Escala de razón	65
4.5.5 Escala Likert	68
4.5.5.1 Análisis de reactivos	72
4.5.5.1.1 Método de correlación	72
4.5.5.1.2 Criterio de consistencia interna	73
4.6 Prueba Piloto	78
4.6.1 Tamaño de la muestra	78
4.7 El instrumento	82
4.7.1 Recomendaciones para la redacción de los reactivos	82
4.7.2 Lista de reactivos	83
4.7.3 Presentación del Instrumento	85
4.7.4 Observaciones	85
4.8 Análisis del instrumento	87
4.8.1 Análisis de reactivos	88
4.8.2 Análisis de confiabilidad	90
4.9 Implementación del cuestionario	92
Conclusiones	94
Glosario	95
Apéndice I (Medidas de correlación)	100
Apéndice II (Teoría de los tests)	119
Bibliografía	133
Direcciones en Internet	135

INTRODUCCIÓN

En la actualidad la tecnología se ha venido incorporando a nuestras vidas; los alcances logrados en campos como la inteligencia artificial, la genética, las telecomunicaciones, etc., han dejado de ser un estímulo para aquellas personas dedicadas a escribir cuentos y libros de ciencia-ficción para convertirse en una realidad latente que de un momento a otro inundará nuestro entorno.

La educación, lejos de intentar desarrollarse de manera independiente a todos estos cambios, debe darse a la tarea de adaptar toda esa serie de descubrimientos y adelantos tecnológicos para su propio beneficio y, de esta forma, diseñar nuevas técnicas que le permitan ser vanguardista en su tarea de enseñar.

En nuestros tiempos las computadoras son una herramienta fundamental para realizar muchas de las tareas que a diario se nos encomiendan, son numerosos los sitios donde se requieren personas instruidas en el uso de estas máquinas, motivo por el cual la demanda por aprender computación se ha venido incrementado últimamente y, ante tal hecho, han surgido diversas escuelas dedicadas a impartir cursos de esta materia; ahora bien, el número de este tipo de escuelas ha aumentado considerablemente durante la última década, esto resulta una verdad innegable, basta con salir un día hacia cualquier sitio para regresar con una cantidad considerable de propaganda (si es que antes no encontraron un buen lugar en algún bote de basura o como adorno en las calles de nuestra hermosa ciudad), sin embargo, la calidad con la que dichos colegios realizan tal labor es cuestionable.

La Dirección General de Servicios de Cómputo Académico (DGSCA) de la UNAM se ha encargado de captar las nuevas tecnologías y adecuarlas a las necesidades de los individuos, ofreciendo a diversos sectores de la población

múltiples y variados servicios, siendo uno de éstos los cursos de computación. La DGSCA busca la calidad en cada uno de los cursos que imparte y una forma de garantizarla es evaluándolos, para de esta forma mejorar los aspectos que así lo requieran, de ahí la importancia de contar con algún instrumento que nos permita llevar a cabo dicha evaluación.

A continuación se listan y describen brevemente los capítulos que componen el presente trabajo:

Primer Capítulo. Se definirán los conceptos de evaluación y medición ya que deberán estar presentes a lo largo de toda la obra.

Segundo Capítulo. Hablaremos acerca de la Educación Continua y describiremos las funciones que realiza la Dirección General de Servicios de Cómputo Académico (DGSCA) dentro ésta área.

Tercer Capítulo. Introduciremos el concepto de Instrumento de Medición y las características que como tal debe poseer.

Cuarto Capítulo. De acuerdo a la teoría existente, se detallarán y llevarán a cabo cada uno de los pasos para construir el instrumento que nos permita evaluar los cursos de cómputo de la DGSCA.

Al final de la obra se encuentran los apéndices en los cuales se desarrolla la teoría matemática sobre la cual esta fundamentada la teoría que aquí se expone.

CAPÍTULO I

LA EVALUACIÓN Y LA MEDICIÓN

1.1 LA EVALUACIÓN.

Para la mayoría de nosotros el escuchar la palabra evaluación significa exámenes o pruebas, ya sean orales o escritas, sin embargo, el concepto que envuelve a dicha palabra va más allá de esta simple idea.

Cuando hablamos de evaluación nos referimos a todas las técnicas que empleamos para determinar hasta que punto los objetivos que nos hemos propuesto (metas) han sido alcanzados, ésto nos da una noción más amplia y vasta de lo que el término significa, ya que en cualquier actividad que emprendamos donde existan objetivos claramente establecidos estaremos en condiciones de aplicar una evaluación, por citar algunos ejemplos podríamos evaluar:

- El trabajo de un empleado en una planta maquiladora.
- La calidad de un producto médico.
- La actuación de una delegación olímpica en una olimpiada.
- El aprendizaje de un alumno al finalizar un curso, etc.

Desde el punto de vista educacional, cabe definir a la **evaluación** como un proceso sistemático para determinar hasta que punto los alumnos alcanzan los objetivos de la educación¹.

La definición anterior para fines educativos es adecuada, pero ahora asumamos el papel de una empresa que ha enviado a un curso de capacitación a sus empleados (empresa A), o el de una empresa dedicada a ofrecer como servicio

¹ Gronlud, Norman Edward "Elaboración de Tests de Aprovechamiento" Ed. Trillas. México, 1974 (Reimp. 1999).

básico cursos, ya sean de manejo, locución, computación, etc., (empresa **B**); en el primer caso la empresa **A** estaría interesada en saber si el curso o programa al que asistieron sus empleados fue una inversión o un gasto, en el segundo caso habría cierta curiosidad por parte de la empresa **B** en saber que tan buenos o malos son los cursos que ofrece. Cualquiera que fuese la situación, es notable que un proceso de evaluación respondería a estas interrogantes.

De la definición de evaluación dada hay dos aspectos importantes y fundamentales que extienden este concepto más allá del ámbito educativo, que son:

1. La evaluación implica un *proceso sistemático* lo cual omite la observación no controlada (al azar), de los factores considerados en dicho proceso.
2. La evaluación siempre supone que los *objetivos* (metas) han quedado previamente identificados, sin éstos, es prácticamente imposible dictaminar el grado de adelanto.

Es de gran importancia mantener siempre presentes estas dos observaciones al momento de llevar a cabo un procedimiento de evaluación.

1.2 LA MEDICIÓN.

La *medición*, en su definición clásica, es el procedimiento mediante el cual asignamos números (calificaciones, medidas) a las propiedades, atributos o características de los objetos (o de unidades experimentales), estableciendo las reglas específicas sobre las que se fundamentan tales asignaciones².

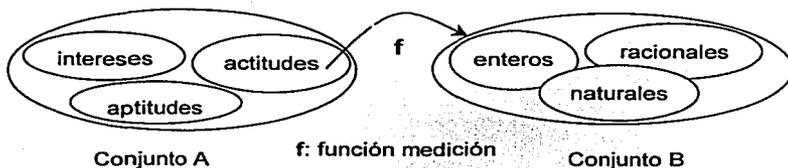
Hay autores que no están muy conformes con el término "objeto", argumentando que no todas las cosas que podemos medir pueden verse o tocarse, y bajo este supuesto dan una nueva definición de lo que significa medir.

² Morales, María Luisa. "Psicometría Aplicada" Ed. Trillas. 2ª Edición. México, 1990 (Reimp. 2002).

Para efectos de nuestro estudio, nos bastará con la definición clásica siempre y cuando consideremos que conceptos como la credibilidad o la inteligencia son medibles pese a no ser tangibles.

De acuerdo con lo anterior tenemos que la medición es una función (regla de correspondencia) que va de un conjunto cuyos elementos son objetos o conceptos abstractos (Conjunto A), a un subconjunto de números reales (Conjunto B). Cuadro 1.1.

Cuadro 1.1



La medición, en contraste con la evaluación, solo se limita al terreno cuantitativo (la asignación de números), la evaluación trasciende al emitir juicios cualitativos, esta es la diferencia primordial entre estos dos conceptos, por tanto podemos decir que la evaluación es un término mucho más amplio e inclusivo que medición.

1.3 CLASIFICACIONES DE LA EVALUACIÓN.

Podemos clasificar a la evaluación por³:

- I. El momento de su aplicación
- II. Lo que pretende medir.

³ Rodríguez, Mauro y Austria, Honorata "Formación de Instructores" Ed. Mc Graw Hill, 2ª Edición, México, 1999.

I. Por el momento de su aplicación.

a) Evaluación diagnóstica o pretest.

Se aplica antes de comenzar un proceso de **enseñanza-aprendizaje**⁴. Los resultados obtenidos servirán como referencia para hacer comparaciones entre el antes y el después.

b) Evaluación formativa.

Se aplica durante el proceso de enseñanza-aprendizaje para determinar en que medida se están alcanzando los objetivos propuestos.

c) Evaluación terminal.

Tiene como propósito evaluar el aprendizaje final, y de acuerdo al tiempo que transcurre entre el momento de realizarla y el fin del curso o programa se clasifica en:

- Inmediata (si se efectúa al concluir el curso o programa).
- A mediano plazo (si se aplica entre 3 y 6 meses después de terminado el curso o programa).
- A largo plazo (si ocurrió después de 6 meses de finalizado el curso o programa).

II. Por lo que pretende medir.

a) Evaluación de la reacción al evento.

Mide el interés del individuo por el acto o evento. En nuestro caso podemos pensar en un curso de capacitación, un diplomado, un seminario, etc.

⁴Ver Glosario página 98.

b) Evaluación del aprendizaje.

Mide el aprendizaje obtenido en el evento.

c) Evaluación del comportamiento.

Mide las capacidades desarrolladas a lo largo del evento, se aplica al final de éste.

d) Evaluación de los resultados.

Mide los logros conseguidos después de aplicar lo aprendido.

1.4 TÉCNICAS DE EVALUACIÓN.

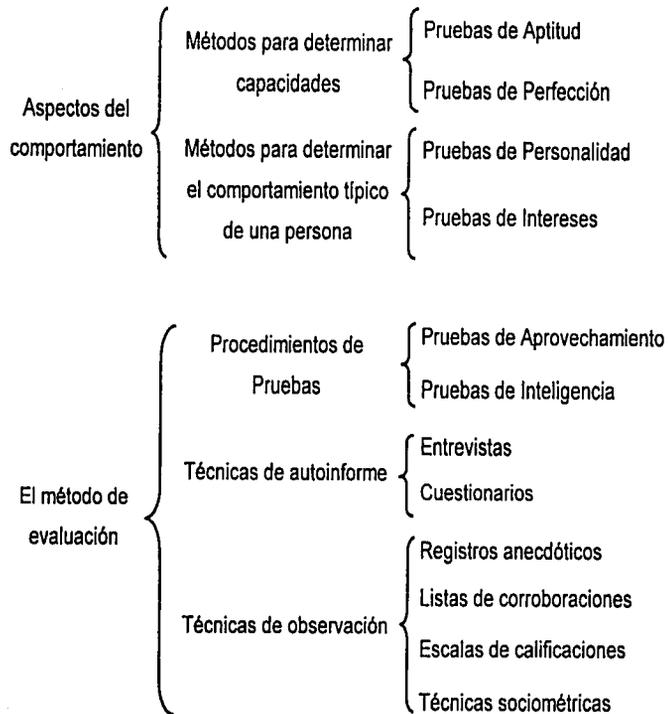
Una característica distintiva del proceso de evaluación es la gran diversidad de métodos a los cuales podemos recurrir para efectuarla.

Es sumamente difícil clasificar los procedimientos de evaluación existentes pues algunos de ellos pueden servir para más de un propósito de acuerdo al contexto en el que se empleen.

Una clasificación algo detallada (pero arbitraria)⁵, de los métodos para evaluar queda expuesta en el cuadro sinóptico: "Clasificación de las Técnicas de Evaluación". Cuadro 1.2.

⁵Gronlund, Norman Edward "Medición y evaluación en la enseñanza". Ed. Trillas. México, 1974.

CUADRO 1.2
CLASIFICACIÓN DE LAS TÉCNICAS DE EVALUACIÓN



De acuerdo con el cuadro sinóptico tenemos que, las técnicas de evaluación, se clasifican en términos del aspecto del comportamiento o en términos del método de evaluación:

I. Aspectos del Comportamiento⁶.

- **Métodos para determinar capacidades.**

a) **Pruebas de aptitud.** Una prueba es meramente una serie de tareas que se usa para medir un ejemplo del comportamiento de una persona en un momento dado. Las pruebas de aptitud son diseñadas para determinar si el individuo tiene facultades o talento latente para aprender determinado oficio, siempre y cuando reciba la instrucción adecuada.

b) **Pruebas de perfección o logro.** Son usadas para establecer el dominio que tiene un individuo en cierta tarea que ha llevado a cabo con anterioridad.

- **Métodos para determinar el comportamiento típico de una persona.**

Este tipo de métodos están conformados por todas las pruebas diseñadas para evaluar intereses, actitudes y aspectos del ajuste personal.

a) **Pruebas de personalidad⁷.** Permiten saber hasta que punto un individuo será capaz de adaptarse a las condiciones del trabajo, en este caso interviene la medición de conceptos como lo son el carácter (ética y moral del individuo) y el temperamento (emociones del individuo).

b) **Pruebas de intereses.** Son utilizadas para determinar los gustos de un individuo hacia cierta actividad. Se esperará que a una persona que le agrade desempeñar algún trabajo lo hará mejor que a la que le desagrada.

⁶ Entiéndase por **comportamiento** el conjunto de actividades que realiza un individuo en una situación determinada.

⁷ La **personalidad** puede definirse como: la organización física y mental en la dinámica del individuo en su esfuerzo por ajustarse a su ambiente.

II. El Método de Evaluación.

- **Procedimientos de Pruebas.** La manera en la que se clasifican las pruebas es muy amplia; las clasificaciones van desde las pruebas orales o escritas, pruebas individuales o de grupo, pruebas de papel y lápiz o instrumentales, y las que se listan a continuación:
 - a) **Las Pruebas de Aprovechamiento.** Esta clase de pruebas son las que comúnmente se emplean en las escuelas y se aplican a los estudiantes, hay varios tipos de clasificaciones aquí se muestran las más comunes.

Clasificación I:

- **Las Pruebas de Ensayo.** La calificación asignada a este tipo de pruebas se ve influenciada por el criterio de quien califica, por tanto son mediciones *subjetivas*.
- **Las Pruebas Objetivas.** Las pruebas objetivas se califican *objetivamente*, lo cual significa que cualquier persona prevista de una clave e instrucción adecuada podrá asignar la calificación.

Clasificación II:

- **Pruebas de Destreza.** Nos permiten establecer si todo participante ha obtenido un mínimo de conocimientos.
 - **Pruebas de Revisión.** Una vez que se asume que todo participante ha alcanzado cierto conocimiento, las pruebas de revisión nos permiten valorar diferencias de conocimiento entre los participantes.
 - **Pruebas de Diagnóstico.** Este tipo de pruebas nos permiten identificar las deficiencias específicas de los participantes.
- b) **Pruebas de Inteligencia.** Describimos este tipo de pruebas al final del Capítulo III.

- **Las Técnicas de Autoinforme.**

- a) **Entrevistas.** Son métodos empleados para obtener información acerca de los intereses y actitudes del individuo por medio de un entrevistador. Un estudio detallado se hará en el **Capítulo IV**.
- b) **Cuestionarios.** Se utilizan para evaluar intereses, actitudes y otros de los aspectos de los ajustes personales y sociales de un individuo. Como en el caso anterior haremos un análisis extenso en el **Capítulo IV**.

- **Técnicas de Observación.**

- a) **Registros Anecdóticos.** Son breves descripciones de algún comportamiento observado en un individuo que pareció importante para propósitos de evaluación.
- b) **Listas de Corroboraciones.** Son listas preparadas de aseveraciones que se refieren a rasgos del comportamiento o actuación en alguna área con respecto a un individuo.
- c) **Escalas de Calificaciones.** Son dispositivos para registrar sistemáticamente las opiniones de los observadores relativas al grado en que se presenta una cualidad o rasgo en un individuo.
- d) **Técnicas Sociométricas.** Es un método para evaluar las relaciones sociales existentes en un grupo de individuos.

Una de las razones por las cuales existen tan variadas técnicas para evaluar es que cada una de ellas proporciona evidencia única pero restringida sobre algún aspecto del cual se pretende obtener información, por eso es recomendable combinar algunas de estas técnicas entre sí para obtener un cuadro más completo de la situación que se examina.

1.5 FUNCIONES DE LA EVALUACIÓN.

A continuación daremos una breve lista de las funciones de la evaluación:

- **El fin primario de la evaluación es orientar.**

La evaluación sirve de guía a un maestro, a sus alumnos, a los padres de éstos últimos, etc., para establecer el grado en el que se están alcanzando los fines educativos. Para una empresa, institución o escuela, evaluar la calidad de sus productos, la funcionalidad de sus procedimientos, las habilidades de sus empleados (para promover ascensos) o la capacidad de los aspirantes a ocupar nuevos puestos, resulta esencial para determinar si los recursos asignados a cada área producen los resultados que se esperan.

- **Toda evaluación cumple hasta cierto punto una función diagnóstica.**

Al valorar, el maestro determina tanto los puntos fuertes como las deficiencias del alumno. Una empresa identifica las anomalías en las tareas que ejerce para corregirlas o ubica las áreas de mayor desempeño para premiarlas.

- **La evaluación tiene una función de pronóstico.**

Mediante la evaluación pueden hacerse conjeturas de manera más efectiva y científica respecto a las posibilidades educativas del alumno en el ambiente académico. Para un proceso de capacitación de los empleados de una compañía, una evaluación apropiada es el comienzo de una planeación, preparación y ejecución de nuevas labores de capacitación.

- **La evaluación constituye un medio de estimular al participante.**

El participante trata de aprender particularmente en los aspectos en los que sabe que será evaluado. Si un empleado es enviado a diversos cursos de capacitación deberá dar su mayor esfuerzo en cada uno de ellos por aprender todo lo necesario, ya que un nuevo conocimiento adquirido que se vea reflejado en las

tareas que realiza en el trabajo y que para la empresa se traduzca en la optimización de servicios, le permitirá mayores aspiraciones labores.

- **La evaluación ofrece medios para que el interesado se juzgue a sí mismo.**

El interesado en progresar profesionalmente utiliza su valoración para determinar sus logros y fallas en las tareas que desarrolla. Una empresa que evalúa los procedimientos con los cuales brinda sus servicios o realiza sus funciones, establecerá en qué medida es eficiente o no lo es.

Quizá existan muchos otros propósitos por los cuales llevar a cabo un proceso de evaluación, como en un principio dijimos, solo hemos expuesto algunos de ellos, al lector le resultará fácil imaginar otros escenarios en donde aplicar los beneficios de una buena evaluación.

Ahora que se han definido los conceptos de evaluación y de medición, se ha mencionado la importancia de la evaluación dentro de diversas áreas y se ha presentado un breve resumen de las técnicas con las cuales se puede evaluar además de citar algunas de las funciones que persigue, hablaremos acerca de las tareas que realiza la Dirección General de Servicios de Cómputo Académico (DGSCA) en la UNAM.

CAPÍTULO II

LA EDUCACIÓN CONTINUA Y LA DGSCA

2.1 LA EDUCACIÓN CONTINUA Y SUS OBJETIVOS.

La **Educación Continua** se identifica como las acciones encaminadas a actualizar, mejorar o profundizar aspectos generales o específicos del área de formación profesional. El objetivo que persigue es complementar, actualizar o perfeccionar aprendizajes en áreas específicas para un eficiente desempeño profesional, personal y social.⁸

Un Centro de Educación Continua se encarga de la coordinación, organización y promoción de programas de educación, como diplomados, cursos y seminarios, dirigidos a ejecutivos, especialistas, empresarios independientes, académicos y a la comunidad en general, para enfrentar los constantes cambios en el entorno. Los programas de Educación Continua proveen de actualización y capacitación en áreas de negocios, tecnología y humanidades, para permitir a las personas adquirir nuevos conocimientos que les permitan ejercer el liderazgo, la dirección efectiva y la realización de sus labores cotidianas de una mejor manera que las lleve al alcance de sus metas.

Con el paso de los años la Educación Continua ha pasado de ser una "modalidad educativa", a representar un modelo de vida: el aprendizaje debe, necesariamente, reciclarse sin cesar, para beneficio del individuo y de su comunidad.⁹

⁸ Tecnológico de Antioquia. Puede consultarse su dirección en Internet al final de estas notas.

⁹ Ibid.

2.1.1 CLASIFICACIÓN DE LOS ACTOS ACADÉMICOS DE LA EDUCACIÓN CONTINUA.

La Asociación Mexicana de Educación Continua (AMEC)¹⁰ ha clasificado los actos académicos que conforman a los centros de educación continua de la siguiente forma:

- Actualización.
- Capacitación y desarrollo.
- Certificación.
- Desarrollo Personal.
- Extensión de la Cultura.

Los cuales se describen a continuación:

• Actualización

Acto académico en que se comunican los enfoques y avances de vanguardia en la disciplina correspondiente.

• Capacitación y Desarrollo.

Acto académico en el que se transmiten conocimientos y se desarrollan habilidades y actitudes que permiten al participante acceder a un entorno laboral o a mantener y/o promoverse en una organización.

¹⁰ La Asociación Mexicana de Educación Continua, A.C. (AMEC) es una organización constituida en 1990, integrada por profesionales comprometidos con el desarrollo del país conscientes de la necesidad de mantenerse a la vanguardia en los adelantos de la educación continua. Se puede consultar su dirección en Internet al final de estas notas.

• **Certificación¹¹**

Acto académico que permite al profesional de un área específica obtener un puntaje para su certificación o prepararse para los exámenes de certificación que le permitan ejercer su profesión óptimamente.

• **Desarrollo Personal**

Acto académico en el que el participante aprende conceptos o desarrolla habilidades que procuran su bienestar y evolución como ser humano.

• **Extensión de la Cultura.**

Acto académico que tiene como objetivo la transmisión de los valores y manifestaciones estéticas.

2.1.2 TIPOS DE ACTOS ACADÉMICOS DE LA EDUCACIÓN CONTINUA.

Dentro de los tipos de actos académicos de la educación continua se encuentran:

• **Círculo de Estudio.**

Esquema de trabajo que se desarrolla durante una o varias sesiones, en las que se reúne un grupo de interesados sobre un asunto determinado con el fin de compartir la información y reflexiones.

Objetivo. Promover el avance en el conocimiento y sobre todo, fomentar el desarrollo de sus participantes.

¹¹ Un título de certificación es un documento que respalda los conocimientos que posee una persona para implementar o dar soporte a cierto producto o tecnología.

- **Conferencia.**

Situación en la que un expositor calificado pronuncia un discurso ante un auditorio.

Objetivo. Presentar información completa y detallada sin interrupciones, identificar problemas y explorar soluciones.

- **Congreso.**

Reunión que persigue el intercambio y la fusión de experiencias y opiniones entre un grupo de personas calificadas en determinado campo.

Objetivo. Actualizar, divulgar y avalar el producto del estudio, la investigación y la experiencia académica.

- **Curso.**

Este es el tipo de acto académico que nos corresponde analizar y que de acuerdo a la AMEC se define como: la experiencia educativa de una o varias sesiones en la que se trabaja básicamente a partir de la exposición y la instrucción magisterial; cuyo objetivo es poner en circulación los conocimientos teóricos o prácticos para que él o los participantes se apropien de ellos.

- **Diplomado.**

Programa curricular que estructura unidades de enseñanza aprendizaje sobre un tópico determinado y que tiene la suficiente extensión y formalidad como para garantizar la adquisición o el desarrollo de un conocimiento teórico y/o práctico, válido.

Objetivo. Que el participante profundice, amplíe, complemente, actualice o en general, que profesionalice su conocimiento y su desempeño en la vida o en el trabajo y que además se le acredite por ello.

• **Foro.**

Sesión o serie de sesiones en el que un grupo de expertos discuten y disertan productivamente sobre un tema.

Objetivo. Avanzar en el conocimiento de un campo disciplinario determinado.

• **Jornada.**

Serie de jornadas concebidas para impartir instrucción e información específica, así como para discutir sobre un tema y llegar a resoluciones y conclusiones comunes.

Objetivo. Identificar, analizar o resolver problemas, para inspirar a la gente hacia la acción, crear conciencia y despertar interés.

• **Mesa Redonda.**

Discusión ante un auditorio por un grupo seleccionado de personas (de 3 a 6) bajo la dirección de un coordinador.

Objetivo. Exponer y enfocar diferentes puntos de vista, hechos y actitudes sobre un tópico, con un máximo de relación e interés.

• **Panel.**

Encuentro de estudiosos con el propósito de presentar ante una audiencia de interesados o ante ellos mismos las distintas visiones sobre un tema determinado.

Objetivo. Al presentar las distintas visiones se busca ampliar perspectivas y referencias en torno a un conocimiento, así como promover la llegada a conclusiones grupales.

• **Seminario.**

Reuniones de trabajo sobre temas de interés, en el que se aportan las distintas perspectivas de los participantes. Se promueve así la mutua enseñanza.

Objetivo. Que cada participante presente un tema y se debata sobre él con el fin de consolidar o producir conocimiento teórico práctico.

• **Simposium.**

Grupo de charlas, discursos o exposiciones presentadas por varios individuos sobre diversas facetas de un solo tema.

Objetivo. Presentar información básica, relativamente completa y sistemática.

• **Taller.**

Sesiones formales de trabajo de equipo de estudiosos para el desarrollo o el análisis de determinado producto material o conceptual, en el campo científico tecnológico o cultural.

Objetivo. Promover el desarrollo de las capacidades del participante y la elaboración de un producto, instrumento o estrategia evaluable, tangible, útil y aplicable.

2.2 LA DIRECCIÓN GENERAL DE SERVICIOS DE CÓMPUTO ACADÉMICO (DGSCA).

2.2.1 Algo de historia.

En octubre de 1981 se creó, por convenio del Rector Dr. Octavio Rivero Serrano, el Programa Universitario de Cómputo, sustentado en la importancia y la trascendencia que el uso de sistemas de cómputo tiene en áreas como la investigación, la docencia, la administración, etc.; cuatro años más tarde, en mayo de 1985, por acuerdo del Rector Dr. Jorge Carpizo, se fundó la Dirección General de Servicios de Cómputo Académico (DGSCA) a partir de los recursos humanos, físicos y financieros con que contaba el Programa Universitario de Cómputo, en ese entonces la DGSCA estaba conformada por la Dirección de Cómputo para la

Investigación. Finalmente, para febrero de 1987, después de ciertas reestructuraciones, la DGSCA pasa a formar parte del subsistema de la Secretaría General de la UNAM, manteniéndose así hasta nuestros días (en el Cuadro 2.1 se muestra la ubicación de la DGSCA dentro de la UNAM).

2.2.2 Funciones de la DGSCA.

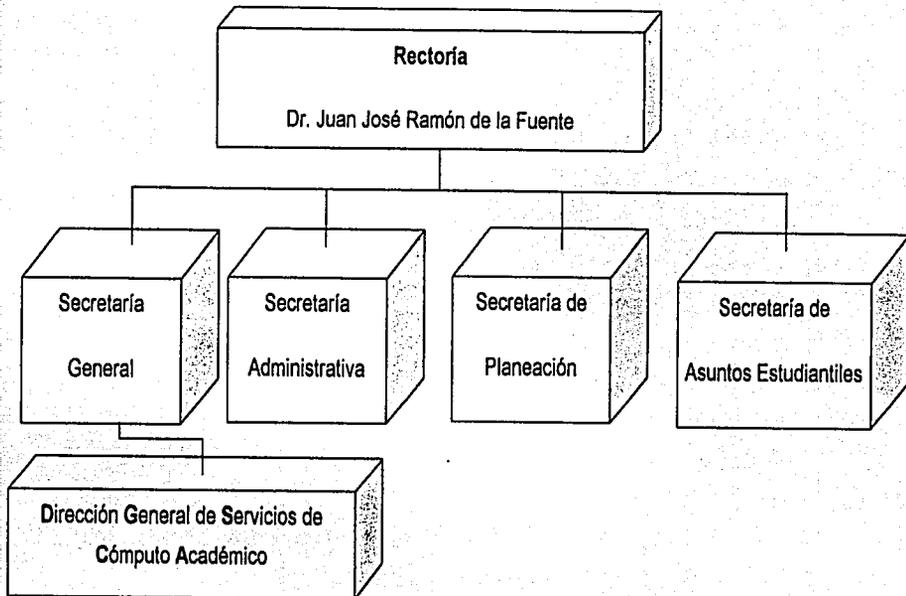
La Dirección General de Servicios de Cómputo Académico (DGSCA) de la UNAM, es la entidad universitaria encargada de supervisar los sistemas centrales de cómputo académico; de operar y extender las telecomunicaciones, del esfuerzo más amplio de capacitación en tecnologías de la información; y de la prospección, innovación y asimilación de estas tecnologías en beneficio de la Universidad y de la sociedad en general. La DGSCA asume el compromiso de absorber nuevas tecnologías y adecuarlas a las necesidades de los individuos, ofreciendo a los diferentes sectores de la población múltiples y variados servicios que giran en torno a los cuatro ejes que fundamentan su existencia¹²:

- Docencia en cómputo.
- Apoyo a la investigación
- Desarrollo de sistemas computacionales.
- Telecomunicaciones.

Nuestro estudio se remitirá al área de Docencia en Cómputo.

¹² Se puede encontrar mas información acerca de la DGSCA en su sitio en Internet, la dirección se encuentra al final de estas notas.

CUADRO 2.1



2.2.2.1 Docencia en Cómputo

En el área de docencia la calidad de los servicios de la DGSCA está avalada por la experiencia de más de cuarenta años en cómputo y tecnología informática¹³, lo cual permite realizar su actividad docente a través de la Educación Presencial y la Educación a Distancia; cada una de ellas se desarrolla con metodologías pedagógicas particulares y responden a necesidades puntuales.

Otras áreas de importancia son¹⁴:

2.2.2.2 Apoyo a la investigación

El supercómputo es la tecnología informática más avanzada que existe para desarrollar investigaciones complejas de alto nivel de especialización; las aplicaciones de la supercomputadora han abierto nuevas líneas de investigaciones científicas en áreas como la ingeniería, la medicina, la geofísica, la geografía, la astronomía, la química, la bioquímica, las ciencias de la atmósfera, las ciencias nucleares y la física, y cada día más disciplinas han ido incorporándose a esta tecnología.

2.2.2.3 Desarrollo en sistemas computacionales

Compuesta por:

- Seguridad en cómputo.
- Sistemas informáticos.

Seguridad en Cómputo

¹³ La UNAM es la primera Universidad en América Latina en instalar una computadora, el lugar donde se llevó a cabo tal evento fue el entonces llamado Centro de Cálculo Electrónico, ubicado en la planta baja del edificio de la Facultad de Ciencias en el año de 1958, siendo Director de esta facultad el Dr. Alberto Barajas Celis. El objetivo principal del centro era realizar investigaciones en las áreas de matemáticas, física y actuaría en las que se requerían grandes cálculos numéricos; a partir de ese momento se marca una nueva era en el desarrollo tecnológico que habría de ser trascendental para la comunidad universitaria y para el país en general.

¹⁴ Tomado de: "Cómputo Académico UNAM. Productos y Servicios". Sin Año. Distribución gratuita.

Uno de los organismos más destacados en materia de seguridad informática en el país y con reconocimiento internacional es el Área de Seguridad en Cómputo (ASC) de la Dirección General de Servicios de Cómputo Académico de la UNAM, esta área está encargada de investigar y diseñar los mecanismos electrónicos que garanticen la integridad de los sistemas de redes de transmisión de datos y los sistemas operativos de las computadoras que opera o tiene a su cargo.

Sistemas informáticos

La DGSCA cuenta con personal altamente especializado para dar asesoría y desarrollar sistemas de información que permitan resolver problemas específicos y automatizar oficinas y áreas críticas.

2.2.2.4 Telecomunicaciones

Conformada por:

- La RedUNAM
- Internet

Red Integral de Telecomunicaciones de la UNAM (RedUNAM)

La RedUNAM es una compleja red de redes, pionera y líder en México. Cumple con la tarea estratégica de comunicar a la comunidad universitaria entre sí y a ésta con el resto de la comunidad académica y científica mundial. Da soporte a más del 90% de las actividades administrativas y culturales de la Universidad Nacional Autónoma de México y permite el intercambio de información entre empresas particulares de todo el país.

Internet

Es un sistema de redes de telecomunicaciones digitales para el acopio y distribución mundial de información que se generan en todo el mundo. Internet es un gran mar de

Uno de los organismos más destacados en materia de seguridad informática en el país y con reconocimiento internacional es el Área de Seguridad en Cómputo (ASC) de la Dirección General de Servicios de Cómputo Académico de la UNAM, esta área está encargada de investigar y diseñar los mecanismos electrónicos que garanticen la integridad de los sistemas de redes de transmisión de datos y los sistemas operativos de las computadoras que opera o tiene a su cargo.

Sistemas informáticos

La DGSCA cuenta con personal altamente especializado para dar asesoría y desarrollar sistemas de información que permitan resolver problemas específicos y automatizar oficinas y áreas críticas.

2.2.2.4 Telecomunicaciones

Conformada por:

- La RedUNAM
- Internet

Red Integral de Telecomunicaciones de la UNAM (RedUNAM)

La RedUNAM es una compleja red de redes, pionera y líder en México. Cumple con la tarea estratégica de comunicar a la comunidad universitaria entre sí y a ésta con el resto de la comunidad académica y científica mundial. Da soporte a más del 90% de las actividades administrativas y culturales de la Universidad Nacional Autónoma de México y permite el intercambio de información entre empresas particulares de todo el país.

Internet

Es un sistema de redes de telecomunicaciones digitales para el acopio y distribución mundial de información que se generan en todo el mundo. Internet es un gran mar de

información diseminada a través de servidores de alta tecnología y sitios electrónicos ubicados en todo el mundo y para todo el mundo.

2.2.3 LA EDUCACIÓN CONTINUA EN LA DGSCA.

La DGSCA, a través de la Dirección de Cómputo para la Docencia, en su papel como institución de Educación Continua, desarrolla programas académicos en cómputo enfocados a distintos sectores tanto de la comunidad universitaria como del público en general, a continuación citamos algunos de ellos.

• Cursos para Empresas e Instituciones

Los programas en este rubro contemplan el diseño de planes específicos de capacitación y actualización en cómputo y telecomunicaciones, acordes con la necesidades e intereses específicos de cada institución o empresa que los solicite.

• Diplomados y Líneas de Especialización.

Diseñados para apoyar a los profesionales de las diversas disciplinas en diferentes áreas del conocimiento asistidos por computadora.

• Programa de Capacitación para el Trabajo

Los programas de esta área se encuentran estructurados en líneas de capacitación en cómputo para la oficina y la edición y administración de la información; esta dirigido a personas con estudios mínimos de secundaria o técnicos.

• Programas de Cursos de Actualización

Conformado por cursos de diversas áreas y aplicaciones de cómputo con el propósito de complementar el desarrollo profesional o académico, este programa está estructurado por 20 áreas académicas que reflejan los últimos avances del cómputo y la tecnología informática.

▪ **Programas de Educación a Distancia**

Mediante la videoconferencia interactiva se realizan cursos, conferencias y diplomados a distancia, usando, como apoyo adicional, Internet y sistemas audiográficos.

▪ **Programa para la Actualización del Personal Docente**

El Programa de Apoyo a la Actualización y Superación del personal docente del bachillerato (PAAS), fue diseñado con el propósito de elevar la calidad académica de los profesores de la Escuela Nacional Preparatoria y del Colegio de Ciencias y Humanidades. La DGSCA participa en este programa impartiendo cursos de computación de diversos niveles para los docentes de los organismos educativos mencionados.

2.2.4 LOS CENTROS DE CÓMPUTO

Los alcances de la tarea docente de la **DGSCA** se amplían y permiten el acceso de otros sectores de la población con la creación de sus **Centros de Extensión**, ubicados en zonas estratégicas de la ciudad de México; de esta manera la Universidad Nacional Autónoma de México contribuye al avance de la sociedad con bases sólidas y en respuesta a las necesidades de la actual demanda.

Los centros con los que actualmente cuenta la DGSCA se listan a continuación:

- Dirección General de Servicios de Cómputo Académico (DGSCA) Ciudad Universitaria.
- Centro de Extensión en Cómputo y Telecomunicaciones Nuevo León.
- Centro de Extensión en Cómputo y Tecnología Informática Máscarones.
- Centro de Extensión en Cómputo y Telecomunicaciones Coapa.

2.2.5 OTROS SERVICIOS DE LA DGSCA

En la **DGSCA** se ofrecen servicios de carácter estratégico en el campo del cómputo y las telecomunicaciones para apoyar las funciones universitarias de investigación, docencia y extensión de la cultura, así como respaldar las actividades que dan soporte al desarrollo de dichas funciones sustantivas.

Cumpliendo con las funciones que tiene asignadas, la **DGSCA** extiende los beneficios de su labor, no sólo a las dependencias de la Universidad que así lo requieren; también se benefician múltiples instituciones de investigación y educación superior externas, así como organismos gubernamentales y privados.

La **DGSCA** ofrece sus servicios en las siguientes áreas:

- **Administración académica e ingeniería de sistemas.**
- **Cómputo de alto rendimiento.**
- **Consultoría en telecomunicaciones.**
- **Laboratorios de extensión.**
- **Multimedios.**
- **Productos interactivos para la docencia.**
- **Prospección tecnológica.**
- **Servicios hemerográficos.**
- **Visualización.**

CAPÍTULO III

INSTRUMENTOS DE MEDICIÓN

3.1 DEFINICIÓN DE INSTRUMENTO DE MEDICIÓN.

Un **Instrumento de Medición** es la herramienta mediante la cual obtenemos información en forma óptima y cuantificable de las variables que son motivo de nuestro estudio.

Ejemplos de instrumentos de medición son: el metro (mide la variable longitud), la báscula (mide la variable peso), el cronómetro (mide la variable tiempo), etc.; sin embargo, y para efectos de nuestro estudio, debemos introducir el concepto de instrumento de medición psicológica.

“Un instrumento de medición psicológica es aquella metodología producida artificialmente, que obedece a reglas específicas y coloca al, o a los individuos, en condiciones experimentales con el fin de extraer el segmento del comportamiento a estudiar y que permita la comparación estadística conductora a la clasificación cualitativa, tipológica o cuantitativa de la característica o características que se están evaluando.”¹⁵

Quando hablamos de Instrumentos de Medición en Psicología, nos referimos a todos aquellos procedimientos u operaciones que permiten llegar a obtener objetivamente y con la mayor certeza posible, información acerca de la expresión de los fenómenos que suceden en esa unidad biológica, social, psicológica, etc., que es la conducta humana.

Hacemos énfasis en el término “Psicológica”, ya que éste marca la diferencia entre los instrumentos que utilizamos para medir características físicas de ciertos

¹⁵ Morales, M. p. 2

objetos¹⁶, de los instrumentos que empleamos para medir determinadas actitudes o comportamientos en los individuos.¹⁷

Los instrumentos de medición psicológica están formados por ítems. "Un ítem es la unidad mínima que compone a una medición, es un reactivo que estimula una respuesta en un sujeto (por ejemplo: una pregunta, una lámina, una frase, un objeto de descripción, una fotografía, etc)"¹⁸. De aquí en adelante los términos ítem y reactivo se usarán de manera indistinta.

El estudio de los instrumentos de medición psicológica cae dentro del área de la métrica mental (psicometría). Ejemplos de este tipo de instrumentos son: las pruebas de inteligencia, las pruebas de aptitudes, los exámenes efectuados en las escuelas (exámenes de álgebra, aritmética, etc.), los tests¹⁹ para medir el IQ (Intelligence Quotient), y en general todas aquellas técnicas para evaluar que consideramos en el Capítulo I.

3.2 ALGUNAS CONSIDERACIONES.

Cuando pensamos en instrumentos de medición como lo son el metro o la báscula, sabemos que el número 0 (cero), es atribuido a la ausencia de longitud o a la carencia de peso; también somos conscientes de que una distancia de 10 km es el doble de una de 5 km o que 250 grs. son la cuarta parte de un 1 kg, sin embargo, los instrumentos de medición psicológica carecen de estas dos características:

¹⁶ Como lo es el "metro", que es utilizado para medir la longitud de algunos objetos físicos.

¹⁷ Los términos "Instrumento de Medición" o simplemente "Instrumento" se usarán de manera indistinta para referirnos a los "Instrumentos de Medición Psicológica o a los "Instrumentos de Medición Física"; según sea el caso, el lector puede inferir fácilmente de que tipo de instrumento hacemos mención ubicando el contexto en el que emplearemos estas ideas.

¹⁸ Hernández, Roberto; Fernández, Carlos y Baptista, Pilar "Metodología de la investigación" Ed. Mc Graw Hill, México, 1996.

¹⁹ Un test es un conjunto de métodos (pruebas, exámenes, experimentos) que permiten valorar o medir una o varias características de un individuo, un grupo, un producto o una máquina.

- Un punto 0 (cero).
- Una escala dividida en segmentos iguales.

Si en un examen de aritmética el alumno Pedro obtiene un cero de calificación, sería arriesgado aventurarnos a afirmar que Pedro nada sabe acerca de operaciones matemáticas elementales, de igual forma si Luis obtiene un nueve de calificación en la misma prueba y Antonio un tres, no estaríamos en condiciones de afirmar que Luis es tres veces mejor que Antonio en aritmética, sin embargo, la falta de un punto cero y una escala con unidades equivalentes en los instrumentos de medición psicológica, no debe conducirnos a la conclusión de que ningún resultado válido ni confiable pueda obtenerse por medio de su aplicación.

En la siguiente sección se hablará acerca de las características que debe poseer un instrumento de medición para poder garantizar que la información que obtengamos mediante su empleo sea válida y confiable.

3.3 CARACTERÍSTICAS DE UN INSTRUMENTO DE MEDICIÓN.

De entre todas las características que debe poseer un instrumento de medición existen dos que son fundamentales: la confiabilidad y la validez.

3.3 La Confiabilidad.

Para ilustrar el concepto de confiabilidad en un instrumento de medición, comenzaremos con algunos ejemplos que nos permitirán darnos una idea de lo que este término significa y posteriormente se dará una definición.

Si el peso de una persona es hoy de 60 Kg. y al usar el día de mañana el mismo instrumento resulta que el peso de esta persona es de 200 kg., todo indica que ese instrumento no es un medio confiable para determinar su peso.

Si de manera independiente dos maestros distintos e igualmente instruidos califican a los mismos alumnos mediante el mismo instrumento y obtienen calificaciones semejantes, podemos decir que los resultados tienen un alto grado de confiabilidad.

Como puede apreciarse en estos ejemplos **la confiabilidad se refiere al grado en que los resultados obtenidos mediante el instrumento se parecen si éste se emplea con los mismos sujetos en repetidas ocasiones.**

3.3.1.1 Factores que influyen en la confiabilidad

Dentro de los factores que influyen de manera directa en cuanto a la disminución de la confiabilidad de un instrumento se encuentran²⁰:

- El calificador.
- El contenido del instrumento.
- El tiempo.

3.3.1.1.1 El calificador.

Vamos a suponer que tenemos dos maestros que son igualmente competentes en una misma materia y a ambos se les han dado los exámenes de cierto grupo para calificarlos. Si las calificaciones otorgadas por los dos profesores son muy diferentes entre sí, estaríamos en condiciones de suponer que nuestro instrumento (el examen), es susceptible al criterio del calificador, en caso de que las calificaciones fuesen semejantes podríamos opinar lo contrario.²¹ Lo que se pretende es que el instrumento al ser evaluado por individuos igualmente capaces nos proporcione

²⁰ Adkins Wood, Dorothy. "Elaboración de Tests. Desarrollo e interpretación de los tests de aprovechamiento" Ed. Trillas. 2a Edición. México, 1990 (Reimp. 2000).

²¹ A fin de realizar convenientemente las comparaciones, hay necesidad de contar con un índice de relación estadística entre dos grupos de calificaciones.

resultados semejantes, es decir, buscamos que la medida obtenida a través del instrumento no se vea afectada por el criterio de la persona que evalúa.

3.3.1.1.2 El contenido del instrumento.

La confiabilidad relacionada con el contenido del instrumento tiene que ver con la elección de una muestra representativa del conjunto de características totales que definen al ente u objeto que deseamos medir. Si nuestras técnicas de muestreo resultan ser inadecuadas, seguramente nuestra confiabilidad se verá disminuida.

Podemos definir un conjunto de características pequeño, con el cual corremos el riesgo de excluir factores relevantes, o por el contrario, puede ser un conjunto grande, más no representativo, si incurrimos en cualquiera de éstos dos casos afectaremos la confiabilidad del instrumento .

3.3.1.1.3 El tiempo.

Aquí hacemos referencia a toda una diversidad de situaciones que pudiesen resultar del hecho de obtener medidas considerablemente distintas si el instrumento es aplicado en dos o más ocasiones (a tiempos distintos), al mismo conjunto de individuos. Si las medidas obtenidas por medio del instrumento serán empleadas como base para hacer predicciones, no deberían variar mucho en cortos intervalos de tiempo.

3.3.1.2 Evaluación de la confiabilidad

Existen varios métodos que nos permiten evaluar la confiabilidad de un instrumento, a continuación citamos los más comunes:

- Método de Test-Retest
- Método de formas equivalentes o paralelas

- Método por mitades (split-half)
- Coeficiente α de Cronbach
- Coeficiente KR20

3.3.1.2.1 Método de Test-Retest.

Este método consiste en aplicar el instrumento de medición 2 veces (a tiempos distintos) a los mismos individuos para posteriormente comparar las medidas obtenidas.

Evaluación: Una vez obtenidas ambas medidas, se procede a calcular el coeficiente de correlación²² entre éstas, si el valor de dicho coeficiente es muy cercano a 1 esto significará que nuestro instrumento posee alta confiabilidad.

Observaciones.

- Se requiere suministrar el instrumento 2 veces, lo cual a veces resulta ser algo complicado.
- La duración del período entre la 1era. y la 2da. aplicación depende de la interpretación que se les dará a los resultados.
- A través de este método podemos determinar que tanto influye el factor tiempo en la confiabilidad de nuestro instrumento.
- Esta clase de evaluación de la confiabilidad no es apropiada cuando tenemos razones para pensar que la primera medición influirá en la segunda. En ocasiones factores como el aburrimiento, la desesperación e incluso el factor

²² La definición de coeficiente de correlación así como la manera en que se calcula se presenta en el Apéndice I.

memoria²³ pueden influir en las medidas obtenidas entre una y otra aplicación, y por consiguiente, en la confiabilidad del instrumento.

3.3.1.2.2 Método de formas equivalentes

En este método se proponen 2 instrumentos que en un principio se presumen equivalentes, (los instrumentos debieron ser contruidos de manera independiente llevando a cabo los mismos procedimientos), ambos son aplicados durante la misma sesión (o en un intervalo de tiempo sumamente corto) a un mismo grupo de individuos.

Evaluación. Una vez obtenidas las medidas de los dos instrumentos, se procede a calcular el coeficiente de correlación entre ambas, la interpretación que se le dará al valor del coeficiente resultante es la misma que en el método anterior (Test-Retest).

Observaciones:

- A diferencia del método anterior (Test-Retest), aquí no nos vemos influenciados por el factor memoria pues pese a que los instrumentos tienen el mismo fin, los elementos (items) que los conforman son distintos.
- Hay que construir 2 instrumentos equivalentes, lo cual en ocasiones es bastante difícil.
- Este método nos permite determinar la consistencia interna del instrumento.

²³ Es muy importante tomar en cuenta el factor memoria en los instrumentos que se aplican en los salones de clase (como lo son los exámenes), ya que a veces el alumno es capaz de recordar las preguntas y simplemente repetir el procedimiento que lo conduzca a la respuesta correcta, sin que lleve a cabo un proceso de razonamiento, que es precisamente lo que en verdad estamos midiendo.

3.3.1.2.3 Método de mitades partidas.

Como su nombre lo dice, este método consiste en dividir el instrumento de medición en dos partes (a la mitad), después de haber obtenido las medidas. Por poner un ejemplo vamos a suponer que tenemos un examen que consta de 10 reactivos, entonces, después de aplicarlo y calificarlo, lo partimos en dos, donde cada una de las partes consta de 5 reactivos (podríamos separar los reactivos de número par de los de número impar) y a cada una de estas partes le corresponde una puntuación.

Evaluación: Por medio de la siguiente fórmula (Spearman - Brown):²⁴

$$\alpha_2 = \frac{2r}{1+r}$$

Donde: α_2 : representa la confiabilidad estimada del instrumento en total.

r: es el coeficiente de correlación calculado a partir de las medidas obtenidas por la aplicación de cada mitad.

Observaciones:

- Si el coeficiente de correlación r es menor a cero, el coeficiente α_2 no podrá ser calculado, esto se debe al hecho de que un valor negativo en nuestro coeficiente r se traduciría en una muy baja consistencia en nuestro instrumento (por no decir nula) y por tanto carecería de sentido calcular α_2 .
- Para el empleo de la fórmula de Spearman-Brown al dividir el instrumento debemos hacerlo de tal forma que el contenido y la dificultad de cada mitad sea equitativa. En caso de que estemos evaluando una sola característica a través de nuestro instrumento, es decir, todos los reactivos tengan el mismo fin, el total

²⁴ Esta fórmula es un caso especial de la "Fórmula Profética de Spearman-Brown" $\alpha_n = (nr) / [1 + (n-1)r]$ para $n=2$ la cual se desarrolla en el Apéndice II.

de formas en las que podemos efectuar la división son las combinaciones de n en $n/2$ divididas por 2, donde n representa el número de ítems del instrumento.

La siguiente expresión nos da el número total de combinaciones:

$$\frac{C_{\frac{n}{2}}^{n/2}}{2} = \frac{n}{2 \left(\frac{n}{2}\right)! \left(\frac{n}{2}\right)!} \quad \forall n \text{ par y } n > 0$$

y por cada división para la que calculemos el coeficiente éste será diferente.

3.3.1.2.4 Coeficiente α de Cronbach

Este procedimiento es usado para determinar la consistencia interna del instrumento; el método para calcular este coeficiente consiste en:

- Aplicar el instrumento.
- Obtener las medidas.
- Emplear alguno de los 2 siguientes procedimientos:²⁵

1) Sobre la varianza de los ítems.

Para lo cual se requiere del uso de la siguiente expresión:

$$\alpha = \frac{n}{n - 1} \left[1 - \frac{\sum_{i=1}^n \sigma^2 (y_i)}{\sigma^2 X} \right]$$

Donde: n : es igual al número de ítems de todo el instrumento.

²⁵ Hernández, R. p. 26

y_i : es igual a la desviación estándar del i -ésimo ítem.

σ^2x : es igual a la varianza de todo el instrumento.

$\Sigma\sigma^2(y_i)$: es igual a la suma de las varianzas de los ítems.

2) Sobre la base de la matriz de correlación de los ítems:

- a) Se calculan los coeficientes de correlación (r) de Pearson entre todos los ítems (todos contra todo de 2 en 2).
- b) Se elabora una matriz de correlación con los coeficientes obtenidos.

La matriz luce de la siguiente forma:

ítems	1	2	3	...	n
1	r_{11}	r_{12}	r_{13}	...	r_{1n}
2	-	r_{22}	r_{23}
3	-	-	r_{33}
:	-	-	-
n	-	-	-	-	r_{nn}

Donde: n : es el número total de ítems de la escala.

r_{ij} : es la medida de correlación r de Pearson del ítem i con el ítem j del instrumento.

Nota:

- La matriz resultante es una matriz simétrica ($r_i = r_{ji}$) y cuadrada de grado n .

- El número de correlaciones que hay que calcular para obtener todos los coeficientes de correlación de la matriz son las combinaciones de n tomadas de 2 en 2:

$$C_n^2 = \frac{n(n-1)}{2}$$

- c) Se calcula el promedio (p) de las correlaciones entre los ítems:

$$p = \frac{\sum \sum r_{ij}}{C_n^2}$$

Donde: C_n^2 : es el número correlaciones descrito anteriormente.

$\sum \sum r_{ij}$: corre para las $i > j$.

- d) Y por último se aplica la siguiente fórmula:

$$\alpha = \frac{np}{1 + p(n-1)}$$

Donde α es el valor del coeficiente.

Observaciones.

- Los valores para α están entre 0 y 1 donde el valor 0 se traduce como "un instrumento muy poco confiable" y el 1 como un "instrumento altamente confiable".
- Este método solo requiere una aplicación del instrumento.

3.3.1.2.5 Coeficiente KR20

Este coeficiente fue desarrollado por Kuder y Richardson y es empleado cuando los ítems de los instrumentos (podemos pensar en preguntas), son dicotómicos, es decir, solo son permitidos dos tipos de respuesta (sí o no, 1 o 0, falso o verdadero, etc.)

Evaluación: A las respuestas obtenidas en los ítems podemos codificarlas mediante el uso de los números 1 y 0 (1 si es verdadero, 0 si es falso o *viceversa*) y después usamos la fórmula KR 20²⁶ definida por:

$$\alpha_2 = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n p_i q_i}{\sigma^2(x)} \right]$$

Donde: α_2 : es la confiabilidad del instrumento en total.

p_i : proporción de las personas que obtuvieron un punto en el i -ésimo ítem del instrumento.

q_i : proporción de las personas que no obtuvieron un punto en el i -ésimo ítem del instrumento ($q_i = 1 - p_i$).

σ^2x : es la varianza total del instrumento.

Observaciones:

- El cálculo de este coeficiente es útil para determinar la confiabilidad en las pruebas académicas ya que podemos asignar un 1 si la respuesta a la pregunta es correcta y un 0 si es incorrecta, sin embargo, no debe emplearse en las pruebas donde la velocidad sea un factor importante.

²⁶Esta fórmula es un caso especial de la fórmula para calcular el α de Cronbach.

- Como la fórmula KR20 es un caso especial del α de Cronbach hereda todas sus propiedades.

La mayoría de las técnicas para obtener los coeficientes de confiabilidad requieren el cálculo de un coeficiente de correlación; como lo sugieren algunos autores²⁷ emplearemos el coeficiente r de Pearson.

3.3.2 La Validez

Pese a que la confiabilidad es una característica con la que deseamos cuente nuestro instrumento, ésta por sí sola no nos proporciona certeza alguna de que la información que obtengamos sea verdadera. Un testigo puede comparecer ante un juez varias veces y emitir el mismo relato de los hechos en cada ocasión, sin embargo, nada garantiza que lo que diga sea verdad, es aquí donde entra el concepto de **validez**.

Decimos que **un instrumento de medición es válido cuando mide lo que se propone medir**, así un metro es apropiado para medir la longitud de un objeto pero no su peso.

Cuando hablamos de validez en instrumentos de medición es conveniente considerar lo siguiente²⁸:

1. "La validez pertenece a los **resultados** de un instrumento de medición y **no** al instrumento mismo. Hablamos a veces de la validez de un instrumento por comodidad pero es más apropiado hablar de la validez de los resultados del instrumento, o más específicamente, de la validez de la interpretación a partir de los resultados."

²⁷ Norville M., Downie "Métodos estadísticos aplicados", Ed. Harla S.A. de C.V. 5ª Edición. México, 1986.

²⁸ Gronlund, N. p. 1

2. "La validez es una cuestión de **grado**. No existe sobre una base de todo o nada. Consiguientemente, debemos evitar pensar sobre los resultados de la evaluación como válidos o no válidos. La mejor manera de considerar la validez es en términos de aquellas categorías que especifican grado como son: gran validez, validez media y poca validez."
3. "La validez es siempre **específica de algún uso particular**. Jamás debe considerarse como cualidad de tipo general. Por ejemplo: los resultados de una prueba de aritmética pueden tener un grado de validez muy elevado ya que indican habilidad para la computación, un grado bajo de validez para medir la capacidad de razonamiento aritmético, un grado intermedio por lo que hace a la predicción de éxito en futuros cursos de matemáticas y cero en validez si se trata de predecir el éxito en el arte o en la música."

3.3.2.1 Evaluación de la validez

Son 3 los métodos principales con los que se juzga la validez y que se usan de manera frecuente²⁹:

- Validez de contenido
- Validez ligada a criterio
- Validez de construcción

3.3.2.1.1 Validez de contenido

Una vez determinada la característica que deseamos medir (por ejemplo: la actitud de los alcohólicos hacia el tratamiento de rehabilitación que reciben en cierta institución, la actitud de los alumnos hacia la Universidad a la que asisten o la actitud de las personas hacia una marca de cerveza, etc.) comienza el proceso de

²⁹ Weiers, Ronald M. "Investigación de Mercados". Ed Prentice Hall- Hispanoamericana S.A. México, 1986.

establecer tanto las variables como los reactivos que nos permitirán hacerlo, usualmente este es un procedimiento exhaustivo y se recurre al criterio de expertos en la materia.

Evaluación. Se compara el contenido del universo por medir con el contenido del instrumento a fin de determinar si la muestra de ítems que compone a éste último es una **muestra representativa** de todos los ítems que se generaron para medir a las variables de estudio, en ocasiones para alcanzar este objetivo es conveniente emplear alguna técnica de muestreo.

3.3.2.1.2 Validez ligada a criterio

Algunas veces deseamos que nuestro instrumento proporcione medidas que nos sirvan para pronosticar el valor futuro de alguna variable o simplemente queremos diseñar un instrumento para medir alguna característica para la cual ya existen mecanismos de medición vigentes pero que resultan ser muy complejos en su modo de operar. Es aquí donde intervienen los métodos de **validación concurrente** y **validación predictiva**.

3.3.2.1.2.1 Validez concurrente

La validez concurrente consiste en comparar la medida conseguida con el instrumento que diseñamos con alguna otra medida existente obtenida mediante algún otro criterio, para tal efecto, esta otra medida debe de haber demostrado ser confiable y válida. En este momento podría surgir la siguiente pregunta: ¿por qué validar un nuevo instrumento si ya hay un método que nos proporciona la información que necesitamos?, bien, la respuesta a esta interrogante es casi siempre la misma, deseamos un ahorro en tiempo y costo tanto en la recolección de las medidas, como en el tratamiento que se les dará a estas para resultados de análisis y toma de decisiones, y una manera de hacerlo es mejorando nuestros procedimientos de medición.

Evaluación. Una vez obtenidas las medidas (la del instrumento y la del otro criterio), se calcula el coeficiente de correlación³⁰ entre ambas, si el valor del coeficiente es cercano a 1 podremos usar indistintamente cualquiera de los dos procedimientos.

3.3.2.1.2.2 Validez predictiva

Este método designa la capacidad que tiene la medición del instrumento de pronosticar algún valor futuro asociado con la variable que pretendemos medir.

Quizá el ejemplo más común de este tipo de validez nos lo da el área de la psicología industrial. La gran mayoría de las "pruebas de aptitud" diseñadas para la selección de personal en las empresas fueron "validadas" correlacionando las calificaciones obtenidas por los trabajadores en las pruebas, con las puntuaciones que les fueron asignadas por el rendimiento en su trabajo, nuevamente pedimos que esta otra medida de comparación sea confiable y válida.

La gráfica A (a menudo llamada "Gráfica de Expectación") resume y compara la información que se recabó mediante la aplicación de un test de aptitudes y las calificaciones que obtuvieron un conjunto de pilotos al final de un curso de pilotaje para un avión³¹. De acuerdo con la calificación que obtenga un aspirante a piloto en el test podremos determinar las probabilidades que tenga de finalizar con éxito su entrenamiento, esto a la larga se convierte en un ahorro para las empresas dada la cantidad de tiempo y esfuerzo que se invierten en cursos de capacitación, para que el empleado pueda realizar su actividad laboral.

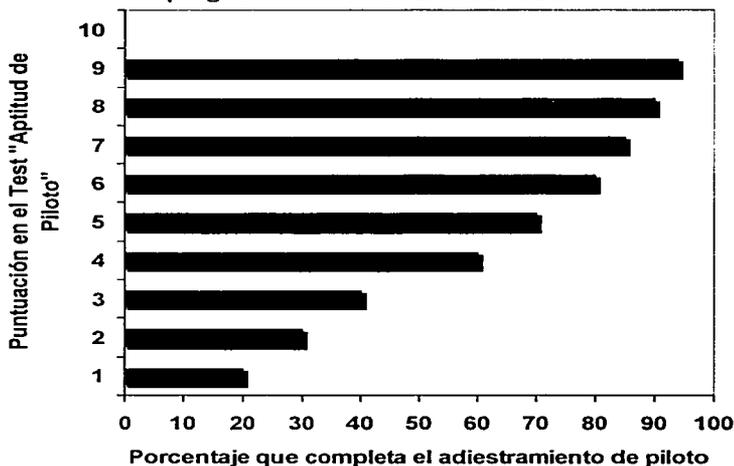
Muchas escuelas, en particular las Universidades, utilizan la validez predictiva para construir sus exámenes de admisión. Se correlacionan las notas de los alumnos que han concluido sus materias con las calificaciones que obtuvieron en el examen de admisión, si la correlación es alta, podremos utilizar las puntuaciones que obtenga

³⁰ Para ver como se calcula el coeficiente de correlación consultar el Apéndice I.

³¹ Dunnette, Marvin "Psicología Industrial" Ed. Trillas. 2ª Edición. México, 1989 (Reimp. 1999).

un aspirante en dicho examen para pronosticar su futuro desempeño a lo largo de sus estudios en cierta institución. La mayoría de las pruebas de aptitud³² son validadas de esta forma.

Gráfica A
Gráfica que muestra la relación entre las puntuaciones que obtuvieron los pilotos en el Test "Aptitud de Piloto" y la finalización con éxito del programa "Adiestramiento de Piloto"



Evaluación. Se correlacionan las medidas del instrumento con alguna otra medida que se obtenga de manera posterior, si el grado de asociación entre ambas variables es alto, estaremos en condiciones de predecir valores a partir de las medidas que arroje el instrumento.

³² Ver Capítulo I página 7

3.3.2.1.3 Validez de constructo

Si en vez de estar interesados en la calificación (medida) que obtiene un alumno en un instrumento que sirve para evaluar sus conocimientos en aritmética estuviéramos interesados en el grado de **capacidad para razonar** del alumno emplearíamos la validez de constructo³³.

La validez de constructo nos ayuda a interpretar los resultados de un test en términos de algún rasgo o cualidad psicológica general.

Algunos ejemplos comunes de elementos de construcción son: la inteligencia, las habilidades para el estudio, la actitud científica, la aptitud para las matemáticas, etc., cada uno de ellos está sustentado en una teoría que puede hacerse actuar en la descripción y predicción del comportamiento de una persona, así, por ejemplo, si decimos que una persona tiene aptitud para las matemáticas, sabremos que esperar de ella en ciertas situaciones específicas.

Evaluación. Usualmente se suele evaluar mediante el procedimiento llamado "Análisis Factorial" para establecer los elementos sobre los cuales recae la variable medida.

El análisis factorial es una técnica estadística que se emplea para resumir la información contenida en un conjunto de variables en función de otro, menor. Las nuevas variables se llaman "factores" o "elementos principales"; el analista trata de identificarlas con los conceptos teóricos que son de interés dentro del contexto de su problema.

La técnica de análisis factorial ha sido empleada por el campo de la psicología en numerosas ocasiones para determinar que factores son los que realmente deberán ser estudiados en un problema determinado. Quizá uno de los estudios más importantes en la historia de la psicología que ha requerido el uso de esta técnica ha

³³ Un **constructo** es una variable medida y que tiene lugar dentro de una teoría o esquema teórico.

sido el desarrollado por Louis L. Thurstone al establecer siete factores básicos, casi independientes, para la medición de la inteligencia y que se listan a continuación³⁴:

- **Aptitud numérica:** Se refiere a la habilidad para hacer cálculos numéricos sencillos con rapidez y exactitud.
- **Aptitud espacial:** (se ha denominado "Factor Espacial"). Consiste en la percepción de relaciones geométricas determinadas y su manipulación por visualización en el espacio.
- **Memoria:** Se refiere a la habilidad que tenga el sujeto para memorizar rápidamente.
- **Comprensión verbal:** Consiste en el conocimiento de vocabulario, habilidad para encontrar analogías verbales y comprensión de lectura.
- **La fluidez verbal:** Hace referencia a la habilidad para expresarse mediante un repertorio adquirido y que puede ser probado a través de discursos preparados o espontáneos.
- **La inducción:** Consiste en la destreza para descubrir una regla o principio básico (como en una serie de números o palabras).
- **Rapidez de Percepción:** Consiste en la percepción rápida y exacta de detalles visuales

Estos factores han servido como base para diseñar los test que miden la inteligencia aunque esta lista no es definitiva ya que actualmente se le han sumado otros factores.

³⁴ Thurstone, Louis L. "La medición de la inteligencia: la aptitud y el interés", Ed. Paidós, Buenos Aires, 1967.

3.4 OTROS ASPECTOS A CONSIDERAR

Aparte de proporcionar resultados que posean un grado satisfactorio de validez y confiabilidad, un instrumento de medición ha de satisfacer ciertos requisitos prácticos:

- Debe ser económico tanto desde el punto de vista del tiempo como del dinero
- Debe ser fácil de administrar y de calificar
- Debe suministrar resultados que el personal con que cuente la institución pueda interpretar y aplicar con precisión.

CAPÍTULO IV

DISEÑO Y CONSTRUCCIÓN DE UN INSTRUMENTO DE MEDICIÓN

4.1 PROCEDIMIENTO PARA CONSTRUIR UN INSTRUMENTO DE MEDICIÓN

Una vez que se ha definido el concepto de instrumento de medición y se han descrito algunas de sus características, toca el turno de describir los pasos a seguir para su construcción, dichos pasos se listan a continuación:

- Definir el objetivo que tendrá el instrumento, es decir, determinar qué es lo que mediremos mediante su empleo.
- Identificar qué instrumento, de los que usualmente se emplean, es el que mejor se adapta a nuestros propósitos.
- Definir las variables que nos permitirán obtener la información que necesitamos.
- Crear los reactivos que nos ayudarán a medir las variables y establecer el nivel de medición (escala) de cada uno de ellos, así como indicar la forma en que se codificarán los datos obtenidos para efectos de análisis.
- Aplicar una prueba piloto del instrumento.
- Sobre los datos recopilados de la prueba piloto, se hacen correcciones al instrumento y también se determina su validez y su confiabilidad.
- Una vez corregido y habiendo demostrado ser válido y confiable para los fines establecidos, habremos concluido con la construcción de nuestro instrumento y estaremos en condiciones propicias para emplearlo.

4.2 OBJETIVO DEL INSTRUMENTO.

La DGSCA, por ser una institución de la UNAM y por la enorme trascendencia que tiene para nuestro país, tiene la responsabilidad y la obligación de mantenerse a la vanguardia en todas las tareas que realiza, y en particular, en la impartición de sus cursos de computación, por éstos y otros motivos que discutiremos más adelante, resulta necesario implementar mecanismos que permitan evaluar de la mejor manera, la situación en la que se encuentran dichos cursos, por tal motivo se considerarán aquellos factores que sean relevantes para su valoración.

Como mencionamos en capítulos anteriores, una forma de determinar que tan eficientes o no, son los cursos de computación que la DGSCA imparte, es por medio de una evaluación, para lo cual construiremos un instrumento de medición cuyo objetivo será:

Evaluar la calidad de los cursos de computación que imparte la Dirección General de Servicios de Cómputo Académico (DGSCA) en la UNAM, para determinar los factores que requieren la aplicación de medidas preventivas y/o correctivas.

Entenderemos por **calidad**³⁵ la conjunción de la eficiencia y la eficacia, ya que la búsqueda de ésta es uno de los retos que toda institución enfrenta en el marco de la globalización que se vive.

4.2.1 Beneficios de una evaluación.

Mediante la evaluación de los cursos, la DGSCA:

- Conocerá el grado en el que se alcanzan los objetivos de cada curso en particular.

³⁵ Esta definición es la misma que emplea la AMEC.

- Se percatará de los puntos fuertes así como las deficiencias de cada curso. Con esta información la institución reconocerá cuáles son los factores que funcionan adecuadamente y también podrá reforzar aquellos puntos débiles tanto como sea necesario.
- Podrá realizar pronósticos acerca de las posibilidades de cada curso. En caso de que un curso tenga alta demanda y obtenga medidas muy altas al momento de ser evaluado, podría propiciar una mayor programación en el calendario de cursos.
- Estimulará al personal docente a superarse. Pese a que hasta este momento no se han determinado de manera explícita los factores del curso que serán evaluados, por citar un ejemplo, podríamos considerar al maestro como un elemento importante (que de hecho lo es, como se verá en secciones posteriores) en el desarrollo de un curso; él, al ser consciente de que será evaluado, fomentará su interés en mejorar en todas las actividades que desempeña dentro del salón de clases y de igual forma, si obtiene alguna puntuación baja en algún factor, tendrá la oportunidad de saber cuál es e intentar mejorar en ese aspecto.

4.3 ¿QUÉ INSTRUMENTO DE MEDICIÓN EMPLEAR?

En la investigación existen dos métodos de recolección de datos que son los más comúnmente usados: la encuesta y la observación, ambos son considerados instrumentos de medición ya que su objetivo esencial es medir variables.

4.3.1 La encuesta.

La encuesta se distingue porque los datos son recopilados mediante preguntas que se les hacen a las personas, ya sea de manera oral o escrita, una lista formal de estas preguntas recibe el nombre de **cuestionario**. Las preguntas se formulan tomando en cuenta la información que deseamos obtener.

De acuerdo al medio que usemos para aplicar la encuesta podemos hacer la siguiente clasificación:

- a) Entrevista personal.
- b) Entrevista por teléfono.
- c) Entrevista por correo.

En la entrevista personal, el investigador obtiene, de manera directa, los datos requeridos del informador.

La entrevista por teléfono es similar a la entrevista personal, salvo por el hecho de que la comunicación entre el entrevistado y el entrevistador es a través del teléfono.

En la entrevista por correo el cuestionario es enviado al informante, para tal efecto se puede recurrir al correo o también el cuestionario se puede hacer llegar anexándolo al producto en cuestión, mediante revistas o con repartidores; una vez que el cuestionario ha llegado a manos del entrevistado, éste lo contesta y lo devuelve empleando el mismo medio que el de envío.

Cabe destacar que en la encuesta el elemento esencial es el cuestionario, ya que de un buen diseño de este, dependerá la calidad de los datos que logremos recopilar.

4.3.2 La observación.

Empleando el sentido más usual, decimos que la observación consiste en el hecho de que una o más personas "observan" lo que ocurre en alguna situación real para clasificar y registrar de manera sistemática, confiable y válida, algún comportamiento o conducta manifiesta.

En la observación los investigadores (observadores) prescinden de las preguntas y optan por seguir los indicios de los objetos o de las acciones que a estos interesan.

Para un estudio mediante observación, el observador debe cumplir con ciertos requisitos y poseer o desarrollar algunas características especiales que le permitan ser lo más objetivo posible al momento de registrar la información, pues de no ser así, los datos obtenidos seguramente estarán altamente sesgados.

El método de observación difiere del método de encuesta por el hecho de que en el primero los sujetos no proporcionan información intencionalmente acerca de sí mismos; en el segundo (método de encuesta), los sujetos ayudan al investigador a reunir los datos, además, el estudio por observación requiere de más tiempo que el estudio por encuesta ya que la recopilación de los datos está basada fundamentalmente en el comportamiento de un individuo, de ahí se deriva que entre mayor sea el tiempo que dedicamos a observarlo mejor será nuestra investigación. En la práctica se ha verificado que tanto el método de observación como el método de cuestionario están expuestos a un considerable margen de error, y que, por lo tanto, deberá disponerse de ellos con sumo cuidado.

Para nuestros fines y por encima del método de observación, se empleará el método de encuesta (uso de un cuestionario). Existen diversas razones, dentro de las que se encuentran:

- El período de duración de un curso en la DGSCA (de 20 a 40 hrs. en promedio). Por la naturaleza del método de observación requeriríamos más tiempo a fin de obtener datos más precisos, sin embargo, usando un cuestionario, la recopilación de los datos sería más rápida y fácil de cuantificar.
- El número de cursos impartidos simultáneamente en las 4 sedes existentes (Centro Coapa, Centro Máscarones, Centro Nuevo León y Ciudad Universitaria). No se cuenta con el material humano suficiente para asignar a cada curso un observador que recopile los datos de interés, en contraste, un cuestionario puede

ser aplicado por una misma persona a un grupo de gente de tamaño considerable.

- Los observadores tendrían que pasar por un proceso de selección, esto último con el fin de elegir y entrenar a la gente adecuada para que los datos recaudados no se vean influenciados por el criterio de quien evalúa, de otro modo, diseñando un buen cuestionario, lo único que tendría que hacer la persona encargada de aplicarlo es recibir un poco de capacitación para dar instrucciones a los respondientes.
- Usualmente un cuestionario permite obtener información de los encuestados que el método de observación difícilmente podría aportar, esto se basa en el hecho de que hay veces en que la mejor manera de obtener información de una persona es preguntarle directamente que opina acerca de un tema determinado.
- Dado que lo que haremos es evaluar un curso de computación, es importante destacar que algunos autores³⁶ señalan al cuestionario como una herramienta útil para obtener datos relacionados con:
 - a) El interés y agrado de los participantes en un curso.
 - b) Las actitudes acerca del instructor, de su experiencia y conducción del curso.
 - c) Las actitudes acerca de la organización, diseño e instalaciones de la institución quien imparte el curso.
 - d) El logro de los objetivos del curso.
 - e) La utilidad que representa a los participantes los temas tratados.
 - f) La utilidad que representa para la empresa la aplicación de lo aprendido.

³⁶ Rodríguez, M. p. 4

Una vez expuestos los motivos por los cuales se ha optado por el método de encuesta examinaremos con un poco más de detalle lo que es un cuestionario, que como antes se dijo, es la parte esencial de la cual está compuesto dicho método.

4.3.3 El Cuestionario.

Un cuestionario es una herramienta que utilizamos para recolectar datos y que consiste en un conjunto de reactivos respecto a una o más variables por medir. Diversos autores coinciden en que para diseñar un muy buen cuestionario no recurrimos a un conjunto de reglas establecidas, ya que de hecho, no existen; lo que se hace es seguir recomendaciones que nos ayudarán tanto en el proceso de diseño como en el de construcción.

El cuestionario (a veces también llamado inventario), se usa comúnmente para evaluar intereses, actitudes y otros ajustes personales y sociales. Cuando el cuestionario es aplicado a un grupo considerable de personas es llamado encuesta.

Un cuestionario tiene diversas ventajas dentro de las que se encuentran:

- Un bajo costo.
- Proporciona los datos requeridos en un tiempo relativamente corto.
- Los datos que se recopilan mediante su aplicación se interpretan con facilidad y se cuantifican fácilmente para el análisis.

4.3.3.1 Tipos de cuestionarios.

Las clasificaciones otorgadas a los diferentes tipos de cuestionarios está basada considerando dos aspectos:

- La estructura.
- El carácter directo.

Un **cuestionario estructurado** es aquel en el que las preguntas siempre siguen un orden establecido, el procedimiento para aplicar el cuestionario debe ser el mismo en cada ocasión, cuando lo anterior no ocurre se dice que tenemos un **cuestionario no estructurado**, es decir, no existe un orden preestablecido para la formulación de preguntas; entre una y otra aplicación del cuestionario hay ciertas diferencias.

Un **cuestionario directo** es aquel en el que el objetivo del cuestionario es totalmente transparente para el entrevistado, es decir, el informante conoce el propósito de la encuesta, en caso contrario, cuando hablamos de un **cuestionario indirecto**, el objetivo de este es disfrazado, lo cual quiere decir que el entrevistado desconoce el motivo de la encuesta; el fin que conlleva esto es que el encuestado proporcione información acerca de temas que le podrían parecer comprometedores (como los relacionados con sus ingresos, temas religiosos, etc.), para lo cual se diseñan preguntas que nos ayudan a sondear una situación y poder sacar conclusiones.

Con las características anteriores podemos hacer combinaciones de tal forma que al final obtenemos 4 tipos diferentes de cuestionarios:

- Cuestionario estructurado directo.³⁷
- Cuestionario no estructurado directo.
- Cuestionario estructurado indirecto.
- Cuestionario no estructurado indirecto.

Describiremos brevemente las características de estos cuestionarios.

³⁷ Algunos autores manejan los términos cuestionario con simulación de objetivo y cuestionario sin simulación de objetivo haciendo alusión al cuestionario indirecto y cuestionario directo respectivamente, en este caso, y como se puede notar, se optó por emplear los dos últimos términos.

4.3.3.1.1 Cuestionario estructurado directo

Este tipo de cuestionario es el más empleado en la investigación. Se caracteriza porque las preguntas que lo componen siguen un orden y el respondiente conoce la finalidad del cuestionario. Todos los cuestionarios se aplican siguiendo el mismo procedimiento, además, se reduce la posibilidad de que el entrevistador influya en la respuesta de los entrevistados y por consiguiente en los resultados, su principal ventaja radica en que las preguntas son fáciles de analizar y codificar.

Desafortunadamente, dado que existe un procedimiento establecido para su aplicación, el cuestionario directo estructurado resulta ser muy poco flexible. Este tipo de cuestionario es conveniente usarlo una vez demostrada su validez y confiabilidad.

4.3.3.1.2 Cuestionario no estructurado directo

En este caso existe una persona encargada de formular las preguntas, sin embargo, el orden de éstas varía entre una entrevista y otra.

Dentro de esta clasificación se encuentran las entrevistas de profundidad, en las cuales existe un entrevistador quien se encarga de elaborar las preguntas al entrevistado según se va desarrollando la entrevista. Este tipo de entrevista permite que el entrevistador profundice en los temas que sean de mayor interés para los fines que a él convengan, por tanto resulta ser muy versátil y flexible, sin embargo, tiene la desventaja de que el criterio del entrevistador se vea reflejado en las respuestas de informante, además, dada la naturaleza de las entrevistas, estas tienden a ser largas y las respuestas deben ser interpretadas por expertos; por consiguiente podrían conseguirse resultados subjetivos, otra desventaja que bien vale la pena mencionar es que las respuestas no pueden ser comparadas entre sí.

4.3.3.1.3 Cuestionario estructurado indirecto

El cuestionario indirecto estructurado tiene la ventaja de que las preguntas están redactadas de tal forma que el respondiente desconoce el propósito que tiene el cuestionario. Esto se hace con el fin de que el informante pueda expresar

realmente sus ideas y puntos de vista con respecto a las variables de estudio; el hecho de que sea estructurado permite que la entrevista sea rápida, poco costosa y la tabulación de los datos sea sencilla.

Algunas de las pruebas de actitudes caen dentro de esta clasificación y están basadas en la teoría de que los conocimientos, la percepción y la memoria de los individuos están condicionados por sus actitudes.

4.3.3.1.4 Cuestionario no estructurado indirecto

Como ya se ha explicado el hecho de que un cuestionario sea indirecto hace que el entrevistado desconozca lo que se está midiendo y esto permite que sus respuestas no se vean influenciadas por alguna razón en particular. Ejemplo de este tipo de cuestionarios son las técnicas proyectivas, como lo son: las asociaciones de palabras³⁸, la complementación de frases y la invención de historias. Las técnicas proyectivas trabajan bajo el supuesto de que el entrevistado, al describir una situación, refleje en ellas algún rasgo que posea.

En las técnicas proyectivas se puede aplicar el mismo estímulo a todos los entrevistados (como desarrollar la misma historia); lo cual permite que las respuestas sean comparables entre sí.

4.3.3.2 Composición de un cuestionario

Independientemente del tipo de cuestionario que se trate, todos ellos poseen los siguientes elementos: instrucciones, preguntas (reactivos) y categorías de respuestas, en algunas ocasiones también se les anexan breves explicaciones del motivo del estudio o también agradecimientos al final.

³⁸ En la asociación de palabras, el entrevistado, después de escuchar una palabra, dice lo primero que se le venga a la mente, en la invención de historias se les muestran cuadros o se les dan ciertas situaciones para que el entrevistado desarrolle una historia.

4.4 DEFINICIÓN DE LAS VARIABLES A EVALUAR

4.4.1 Tipos de variables

En la mayoría de los trabajos de investigación las variables que pretendemos medir caen dentro de alguna de las siguientes categorías³⁹:

- Variables del estado mental.
- Variables de estado.
- Variables conductuales.

Variables del estado mental. Son variables internas del individuo a quien estamos estudiando. A esta categoría pertenecen las actitudes y las características de la personalidad, entre otras. En la mayoría de los casos este tipo de variables resultan ser difíciles de medir debido a su carácter interno.

Variables de estado. Son variables externas del individuo y son más fáciles de medir y cuantificar que las anteriores, ejemplo de este tipo de variables son la edad, los ingresos, el sexo, el nivel de escolaridad, etc.

Variables conductuales. Son variables que denotan las acciones pasadas, presentes o futuras de un individuo.

4.4.1.1 Aspectos por considerar.

Para evaluar los cursos de cómputo que la DGSCA imparte remitiremos nuestro estudio a aquellos factores que los conforman, motivo por el cual analizaremos el **proceso educativo** que se lleva a cabo dentro de un salón de clases

³⁹ Weiers, R. p. 38

y en el cual intervienen múltiples variables. Dentro del proceso educativo existen diversos elementos, de los cuales identificamos:

- La elaboración de planes y programas de estudio.
- La aplicación de métodos de enseñanza.
- El uso adecuado de materiales didácticos.
- La calidad de la enseñanza.
- El aprovechamiento de la infraestructura académica.

Los componentes arriba citados dependen, en la mayoría de los casos, del personal docente; debido a que las habilidades que posea para motivar el aprendizaje de sus alumnos, la planeación y conducción de sus clases y su desempeño académico, entre otras cosas, serán piezas fundamentales del buen funcionamiento del proceso educativo. A pesar de que se cuente con excelentes planes de estudio, materiales didácticos de vanguardia e incluso los más modernos métodos de enseñanza, inevitablemente todo sistema educativo sigue dependiendo de la calidad y trabajo de los profesores, por tal motivo resulta de vital importancia evaluar la labor docente; sin embargo, llevar a cabo esta última tarea no es algo sencillo dada la existencia de múltiples variables que debemos examinar, no obstante, las investigaciones han demostrado que, para valorar a los instructores, hay por lo menos 3 aspectos que considerar:

- a) Los indicadores, elementos y criterios que determinan una buena enseñanza o un buen profesor.
- b) Los procedimientos e instrumentos que establecen como se llevará a cabo la evaluación.
- c) Los agentes evaluadores que proporcionen información confiable y certera sobre la calidad de la enseñanza.

4.4.1.1.1 Indicadores de una buena enseñanza

Con respecto a este punto las investigaciones señalan que los indicadores de la enseñanza más utilizados son:

- Estrategias de exposición clara.
- Presentación del propósito de la clase.
- Introducciones persuasivas.
- Uso de ilustraciones y ejemplos de los conceptos principales.
- Uso de estrategias formativas.
- Uso de auxiliares de apoyo al aprendizaje.
- Conductas del profesor que reflejan entusiasmo e interacción positiva con los estudiantes.
- Participación de los alumnos.
- Supervisión del trabajo académico.
- Procedimientos de evaluación del aprendizaje.
- Cantidad de aprendizaje logrado en el curso.⁴⁰
- La satisfacción de los alumnos sobre la calidad del curso.

⁴⁰ Algunos autores sugieren que el nivel de aprendizaje percibido por los alumnos en un curso podría usarse como predictor de la calidad del mismo.

En un estudio realizado en México los indicadores que obtuvieron un peso más alto son⁴¹:

- Preparación del curso.
- Estrategias de exposición y conducción de la clase.
- Dominio de conocimientos sobre la materia.

En la Facultad de Psicología de la UNAM para evaluar al personal docente se estudiaron los siguientes factores⁴²:

- Cumplimiento.
- Relación maestro alumno.
- Calidad de la exposición.
- Capacidad académica.
- Método de trabajo.
- Sistema de evaluación.
- Bibliografía revisada en el curso.
- Programa de la materia.
- Relaciones externas de la materia.

En el mismo estudio se empleó la técnica de análisis factorial⁴³ y se

⁴¹ Martínez-Guerrero, J. & Sánchez-Sosa, J. J. "Intervención social de estrategias docentes en la Facultad de psicología. Métodos Docentes, UNAM 3, México, 1981.

⁴² Ortega, P. "La evaluación de la actividad docente en el análisis de la formación profesional". En J. Urbina, Ed. El Psicólogo: Formación Ejercicio Profesional y Prospectiva. UNAM. México, 1989.

⁴³ Ver Capítulo III página 42.

identificaron 5 factores que bien pueden servirnos de guía para evaluar el desempeño docente:

- Actitud general hacia el curso.
- Actitud hacia los exámenes.
- Actitud hacia el método.
- Relación formador-estudiante.
- Actitud hacia la carga de trabajo.

Una actitud es un síntoma, señal o manifestación relativamente estable de respuestas consistentes dadas por un individuo con respecto a un objeto psicológico ya sea símbolo, slogan⁴⁴, producto, institución, etc; o dicho de otra manera, una actitud es un estado mental que predispone a un individuo a responder de cierta manera cuando está sometido a determinado estímulo.

Las actitudes poseen dos propiedades: dirección e intensidad. La dirección nos indica si nuestra actitud hacia cierto estímulo u objeto es positiva o negativa (favorable o desfavorable) y la intensidad nos indica en que grado lo es (muy agradable, poco agradable, nada agradable, etc.).

Muchos de nosotros hemos desarrollado ciertas actitudes hacia muchas situaciones, objetos o personas de nuestra vida diaria; a algunas personas nos apasiona el fútbol en la misma medida en que a otras les agrada el patinaje artístico sobre hielo, por tanto podemos decir que la actitud desarrollada hacia tales estímulos es favorable, pero también tenemos el caso opuesto, en que objetos como el cigarro o los anuncios ofreciendo bebidas alcohólicas pueden parecer desagradables a cierto grupo de gente, motivo por el cual concluimos que su actitud hacia estos últimos es desagradable.

⁴⁴ El slogan es una frase que atrae la atención del público para difundir una idea o un producto.

Las actitudes, de acuerdo a su definición, pertenecen a las variables del estado mental, las cuales describimos en párrafos anteriores.

Como bien se puede apreciar, la mayoría de los indicadores que se han expuesto centran su atención no solo en el factor personal docente sino también en aquellos elementos que interactúan junto con él dentro del salón de clase, como lo son: el programa, los métodos didácticos, las actitudes hacia la carga de trabajo (tareas) etc., por tanto, no estamos restringiendo nuestra evaluación al desempeño de los profesores sino a todos los factores que componen un curso en sí; cabe destacar que la mayoría de los indicadores citados en nuestra lista han sido validados socialmente y han orientado el diseño y construcción de instrumentos de medición para evaluar la calidad del personal docente y por consiguiente de los cursos que imparten.

4.4.1.1.2 Procedimientos de evaluación

Como se ha venido mencionando, el procedimiento a través del cual evaluaremos los cursos de cómputo de la DGSCA será un cuestionario.

4.4.1.1.3 Agentes evaluadores

Las fuentes a las que se ha recurrido para evaluar la labor del personal docente son muy variadas, pero de entre todas ellas podemos citar:

- Los cambios en actitudes, percepciones y conductas de estudiantes.
- Las autoevaluaciones de profesores.
- Las evaluaciones de jefes de departamento o de comisiones académicas.
- Los registros de conductas docentes realizados por observadores capacitados.
- Las evaluaciones de los alumnos (estudiantes) sobre la calidad docente.

Los estudios demuestran que de todas las fuentes listadas los datos más válidos se han obtenido de las evaluaciones hechas por los alumnos. Investigaciones recientes demuestran que las calificaciones que los estudiantes asignan a la calidad de un curso escolar poseen un alto grado de validez y confiabilidad, razón por la cual los investigadores emplean los resultados obtenidos de los educandos como fuente útil para mejorar la enseñanza⁴⁵, ahora bien, si las evaluaciones de los estudiantes reflejan adecuadamente su opinión acerca de la calidad de la enseñanza que reciben, esto no necesariamente significa que los resultados que se obtengan sean medidas directamente relacionadas con todos los factores que determinan un buen curso.

4.4.1.2 Establecimiento y alcance de las variables por evaluar.

Basándonos en los estudios relacionados con los indicadores que hasta el momento se han expuesto, enseguida se listan las variables que evaluaremos junto con una breve descripción del alcance que tendrán.

- **Dominio de conocimientos sobre la materia**

Conocimientos del maestro con respecto a los temas del programa, profundidad y precisión en las explicaciones dadas a los alumnos y claridad en las respuestas que surjan de sus dudas.

- **Estrategias de exposición y conducción de la clase**

Secuencia en las ideas, énfasis en puntos importantes, uso de diversos métodos didácticos, ejemplos claros, cargas y ritmos de trabajo.

- **Relación del profesor con los alumnos**

Actitud del profesor hacia los alumnos y métodos de participación.

⁴⁵ Una técnica muy usada para perfeccionar la calidad de la enseñanza es la realimentación.

▪ **Cumplimiento docente-institucional**

Cumplimiento del programa oficial del curso y logro de los objetivos propuestos.

▪ **Procedimiento para evaluar el aprendizaje de los alumnos**

Actitud hacia las formas de evaluación, congruencia entre las evaluaciones y lo visto en clase, análisis y entrega de resultados.

▪ **Empleo de auxiliares didácticos**

Uso de pizarrón, computadora, materiales ilustrativos, textos o notas de apoyo, auxilios audiovisuales, etc.

▪ **Infraestructura**

Buen funcionamiento de las instalaciones y el equipo con el que la institución debe contar para llevar a cabo su labor.

Serán estas variables las que determinarán los reactivos con los cuales construiremos el cuestionario para evaluar la calidad de los cursos de la DGSCA.

4.5 DEFINICIÓN DE LOS REACTIVOS Y DE LA ESCALA DE MEDICIÓN.

En este paso además de definir los ítems o reactivos que medirán las variables que hemos establecido y que conformarán el cuestionario, se establece el nivel de medición de cada uno de ellos; en general hay cuatro formas a saber: nivel de medición nominal, nivel de medición ordinal, nivel de medición por intervalo y nivel de medición de razón; estos niveles de medición también son conocidos como escalas. Cuando hablamos de una **escala** aludimos a una serie numérica que muestra los diferentes valores que puede tener una variable, así que, de ahora en adelante, escala o nivel de medición tendrán el mismo significado.

Antes de dar a conocer los reactivos de nuestro instrumento revisaremos brevemente en que consiste cada una de las escalas mencionadas.

4.5.1 Escala nominal

En este tipo de escala simplemente asignamos símbolos para distinguir entre diferentes categorías, los símbolos no se deben manipular aritméticamente, ejemplo de variables que podemos medir mediante una escala nominal son: el *sexo* (variable dicotómica pues asume solo dos valores), tipo de *escuela a la que asiste un estudiante* (pública, privada, etc.).

Como medios descriptivos de información, para la escala nominal, podemos emplear **frecuencias o porcentajes**.

4.5.2 Escala ordinal

En la escala ordinal se tienen diversas categorías, pero además estas mantienen un orden de mayor a menor (*o viceversa*). Los símbolos o etiquetas que asignamos a las clases si indican grado, sin embargo, las distancias que existen entre los grados no es la misma.

Si usáramos una escala ordinal para ponderar la posición de los diferentes puestos de la Empresa A podríamos hacerlo de alguna de las formas expuestas en los Cuadros 4.1 y 4.2.

Para este ejemplo conforme más alto sea el número que posea una persona mayor será el puesto que ocupe dentro de la empresa. Los Cuadros 4.1 y 4.2 nos permiten ejemplificar que no importa cual sea el valor que asignemos a las categorías ya que estos únicamente nos sirven para establecer un orden, así que mientras estemos trabajando con escalas ordinales nos resulta indistinto usar cualquiera de los dos cuadros.

Cuadro 4.1

Empresa A	
Puesto	Jerarquía (Valor)
Presidente	10
Vicepresidente	9
Director	8
Subdirector	7
Gerente	6
Subgerente	5
Empleado	4
Intendencia	3

Cuadro 4.2

Empresa A	
Puesto	Jerarquía (Valor)
Presidente	987
Vicepresidente	7
Director	5
Subdirector	4
Gerente	-3
Subgerente	-5
Empleado	-100
Intendencia	-9184

En la escala ordinal se aplican medidas de descripción estadística como lo son: la **moda**, la **mediana** y los **percentiles**⁴⁶.

4.5.3 Escala de intervalo

En esta escala existen categorías y un orden o jerarquía entre ellas, pero además, se asume que los intervalos entre una y otra categoría son iguales, existe el número cero, pero su asignación es arbitraria.

Uno de los ejemplos más comunes de este tipo de escala viene dado por la medición de la temperatura que se hace empleando la escala en grados Celsius o la

⁴⁶ **Percentil.** Punto de distribución por encima y por debajo del cual se sitúa un determinado por ciento del grupo. Los 99 puntos percentiles dividen una distribución en 100 partes, conteniendo cada una de ellas el 1% de los casos.

escala en grados Fahrenheit, el cero es arbitrario, e incluso para ambas escalas es diferente.

La existencia de un cero arbitrario impide que se puedan dar proporciones entre las medidas asignadas, lo cual quiere decir que si en la región A se encuentra una temperatura de 15° F y en la Región B una temperatura de 30° F no podemos concluir que en la región B la temperatura es el doble que en la región A.

Este nivel permite la aplicación de algunas operaciones aritméticas elementales (suma, resta, multiplicación y división) y por ende trabajar con algunas medidas estadísticas como lo son la **media** y la **varianza**.

Las mediciones del comportamiento humano como lo son las actitudes, las aptitudes, las pruebas de inteligencia, etc. se cuantifican mediante el uso de las escalas de intervalo.

4.5.4 Escala de razón

Este nivel (o escala), posee las mismas características que los niveles de medición anteriores (nivel de medición nominal, ordinal y de intervalo), sin embargo, el cero en este escala no es arbitrario, cuando asignamos el cero a una variable empleando este nivel de medición estamos indicando la carencia de alguna propiedad.

Ejemplos de variables que podemos medir usando la escala de razón son:

- Los ingresos de una persona
- El número de focos que hay en una casa
- El número de hijos por familia
- El número de goles a favor que lleva un equipo de fútbol en un torneo, etc.

Es importante mencionar que una variable puede ser valorada con diferentes escalas de acuerdo al propósito de medición. Como lo muestra el siguiente ejemplo (Cuadro 4.3), la variable: *kilometraje de un vehículo*, es medida empleando dos tipos diferentes de escalas:

CUADRO 4.3

Escala	Categoría
Ordinal	Poco kilometraje (0 a 50,000 km) Mediano kilometraje [50,000 a 300,000 km] Mucho kilometraje [300,000 en adelante]
Razón	Número de kilómetros recorridos: X

El Cuadro 4.4 es un resumen de las escalas que hemos expuesto junto con un ejemplo, una pequeña descripción y los medios que, de acuerdo a la escala elegida, podemos emplear para realizar un análisis con los datos recabados.

Para la construcción de nuestro instrumento de medición se ha optado por emplear una escala de intervalo, el motivo es simple: de acuerdo con el Cuadro 4.4 en las escalas de razón y de intervalo es donde se pueden aplicar más métodos de análisis, en particular, el de correlación, que es el método esencial sobre el cual descansan la gran mayoría de las fórmulas que usamos para determinar la validez y la confiabilidad de un instrumento; la diferencia principal entre estas dos escalas es el punto cero, que como bien se aprecia, en ambas existe, sin embargo, para la escala de intervalo la manera en la que se establece es arbitraria.

CUADRO 4.4

Escala de Medición			
Escala	Ejemplo	Medios para	
		Descripción	Análisis
<p>Nominal</p> <p>Solo se emplea para distinguir entre categorías.</p>	<p>Sexo: F ó M.</p> <p>¿Posee automóvil?: 1= si, 2= no.</p> <p>Estado Civil: <u>S</u>oltero, <u>C</u>asado, <u>O</u>tro.</p>	<p>Frecuencia.</p> <p>Porcentaje.</p> <p>Moda.</p> <p>Tabulación Cruzada.</p>	<p>Prueba Ji Cuadrada.</p>
<p>Ordinal</p> <p>Existe una jerarquía entre los valores o símbolos asignados.</p>	<p>Nivel de escolaridad.</p> <p>Nivel de preferencia de la cerveza X con respecto a las cervezas W y Z.</p>	<p>Mediana.</p> <p>Moda.</p> <p>Percentil.</p>	<p>Correlación de rango-orden.</p> <p>Prueba del signo.</p>
<p>Intervalo</p> <p>Hay jerarquía, se asumen intervalos iguales, existe un punto cero pero es arbitrario.</p>	<p>Escala Fahrenheit.</p> <p>Escala Celsius.</p> <p>Escala de Mercalli.</p>	<p>Media.</p> <p>Desviación estándar.</p> <p>Correlación Pearson</p> <p>Varianza</p>	<p>Análisis de correlación.</p> <p>Análisis discriminatorio.</p> <p>Análisis de varianza.</p> <p>Escala multidimensionales métricas.</p>
<p>Razón</p> <p>Posee todas las propiedades anteriores más un punto cero absoluto.</p>	<p>Edad.</p> <p>Ingresos.</p> <p>Número de clientes de la empresa X.</p>	<p>Media.</p> <p>Desviación estándar.</p> <p>Correlación Pearson</p> <p>Varianza</p>	<p>Análisis de correlación.</p> <p>Análisis discriminatorio.</p> <p>Análisis de varianza.</p> <p>Escala multidimensionales métricas.</p>

Dado que lo que haremos es evaluar un curso de computación, y en vista de las variables que se han definido, podemos concluir que trabajamos con variables del estado mental (o variables psicológicas) y para medir estas variables se construirá un instrumento de medición psicológica, que como se explicó en el Capítulo III; este tipo de instrumentos carecen de un punto cero y de intervalos iguales⁴⁷, aunque como exponen algunos autores el problema de construir una escala de intervalo para variables psicológicas esta resuelto haciendo ciertas consideraciones⁴⁸, por tanto, solo resta el problema del punto cero, el cual podemos fijarlo arbitrariamente, y a la luz de este razonamiento caemos en cuenta de que estamos hablando de una escala de intervalo.

Una vez determinada la escala que usaremos, toca el turno de describir la escala Likert, que como ya ha de suponerse es una escala de intervalo y es la comúnmente empleada cuando recurrimos a esta clase de escalas.

4.5.5 Escala LIKERT

La escala Likert es, en sentido estricto, una escala ordinal, sin embargo, puede operarse como si se tratara de una escala de intervalo. Podría suponerse que el proceso de construir una escala Likert resulta ser relativamente sencillo, el primer paso consiste en generar un conjunto de reactivos, en este caso son afirmaciones que reflejan una actitud hacia la variable que pretendemos medir, el conjunto de afirmaciones que se obtiene es lo que hemos venido llamando instrumento de medición. A cada afirmación le corresponde la misma categoría de respuestas y éstas deberán estar equilibradas, es decir, habrá igual número de opciones favorables como desfavorables y se les asignará un valor para efectos de codificación.

⁴⁷ Ver Capítulo III páginas 26 y 27.

⁴⁸ Magnusson, David "Teoría de los tests: Psicología diferencial. Psicología aplicada. Orientación Vocacional". Ed. Trillas. 2ª Edición. México, 1990 (Reimp. 2000).

Cada sujeto a quien se le aplique el instrumento deberá contestar cada una de las afirmaciones eligiendo la respuesta que mejor exprese su sentir. La medida final se obtendrá sumando las puntuaciones parciales que el sujeto haya dado a cada reactivo. Debido a esta última característica la escala Likert es llamada también "Escala Aditiva". Veamos un ejemplo.

Supongamos que deseamos medir la actitud de cierto sector de una población hacia el fútbol, en este caso emplearemos 2 afirmaciones, las cuales se muestran a continuación, con sus respectivas categorías de respuestas:

Afirmación 1: Es aburrido ver un juego de fútbol.

Categoría de Respuestas:	Totalmente de acuerdo..... (5)
	De acuerdo (4)
	Neutral (3)
	En desacuerdo..... (2)
	Totalmente en desacuerdo (1)
Codificación:	_____↑

Afirmación 2: El fútbol está próximo a ser el mejor de los espectáculos deportivos.

Categoría de Respuestas:	Totalmente de acuerdo..... (5)
	De acuerdo (4)
	Neutral (3)
	En desacuerdo..... (2)
	Totalmente en desacuerdo (1)
Codificación:	_____↑

Como se puede apreciar cada una de las afirmaciones viene acompañada de un conjunto de respuestas de las cuales el respondiente tendrá que marcar la que mejor exprese sus preferencias; si él elige más de una se anula el reactivo. Típicamente son 5 las opciones que se dan como alternativas de respuesta, aunque este número puede variar dependiendo de la capacidad de discriminación que tengan las personas a quien se les aplique el instrumento (siempre y cuando el número de alternativas de respuestas no sean menor a 3); cabe destacar que **las categorías de respuestas deberán ser las mismas para cada reactivo que conforme el instrumento.**

Para este ejemplo las categorías de respuestas son:

- a) Totalmente de acuerdo
- b) De acuerdo
- c) En duda
- d) En desacuerdo
- e) Totalmente en desacuerdo

aunque existen variantes de estas como lo son:

- a) Completamente de acuerdo
- b) De acuerdo
- c) Ni de acuerdo ni en desacuerdo
- d) En desacuerdo
- e) Completamente en desacuerdo, etc.

La opción de en medio (En duda, Ni de acuerdo ni en desacuerdo, etc.) a veces es eliminada del instrumento final ya que los investigadores no han logrado

llegar a un consenso sobre la inclusión de esta opción, algunos de ellos argumentan que todo individuo posee una actitud (ya sea favorable o desfavorable) hacia cierto objeto o fenómeno y por tanto debe expresarla; también hay ocasiones en las que hay cierta tendencia de los respondientes a inclinarse por esta opción, en caso de que esto suceda lo conveniente sería desecharla de las categorías, pues no nos proporcionaría nada de información⁴⁹.

En el ejemplo anterior los números que vienen dentro de los paréntesis nos permitirán, una vez aplicado el instrumento, codificar las respuestas, no hay necesidad de mostrarlos a los respondientes.

La codificación consiste en asignar un valor numérico que represente a la respuesta y que nos permita efectuar análisis estadístico. En general se recomienda asignar el número mayor a la respuesta más favorable y el menor a la respuesta más desfavorable siempre y cuando la afirmación sea positiva, en caso contrario, es decir, que la afirmación sea negativa, se recomienda hacerlo a la inversa. Los números asignados pueden ser cualesquiera, ya que como antes se dijo, la escala Likert es una escala ordinal, la suposición que hacemos es que entre cada una de las categorías de respuestas la distancia es la misma, por tanto, estamos asumiendo que trabajamos con una escala de intervalo.

Resulta de suma importancia destacar que los reactivos usados en la escala de Likert no son sólo afirmaciones sino que se pueden extender a preguntas u observaciones⁵⁰. La escala de Likert es empleada, en la mayoría de los casos, para medir actitudes con respecto a algún concepto o fenómeno, por ejemplo: el presidente de una nación, en otras ocasiones lo que nos interesa es medir los atributos de un objeto, por ejemplo: la dulzura de un caramelo, la nitidez de una imagen, lo confortable de un asiento, etc.

⁴⁹ Weiers, R. p. 38

⁵⁰ Hernández, R. p. 26

llegar a un consenso sobre la inclusión de esta opción, algunos de ellos argumentan que todo individuo posee una actitud (ya sea favorable o desfavorable) hacia cierto objeto o fenómeno y por tanto debe expresarla; también hay ocasiones en las que hay cierta tendencia de los respondientes a inclinarse por esta opción, en caso de que esto suceda lo conveniente sería desecharla de las categorías, pues no nos proporcionaría nada de información⁴⁹.

En el ejemplo anterior los números que vienen dentro de los paréntesis nos permitirán, una vez aplicado el instrumento, codificar las respuestas, no hay necesidad de mostrarlos a los respondientes.

La codificación consiste en asignar un valor numérico que represente a la respuesta y que nos permita efectuar análisis estadístico. En general se recomienda asignar el número mayor a la respuesta más favorable y el menor a la respuesta más desfavorable siempre y cuando la afirmación sea positiva, en caso contrario, es decir, que la afirmación sea negativa, se recomienda hacerlo a la inversa. Los números asignados pueden ser cualesquiera, ya que como antes se dijo, la escala Likert es una escala ordinal, la suposición que hacemos es que entre cada una de las categorías de respuestas la distancia es la misma, por tanto, estamos asumiendo que trabajamos con una escala de intervalo.

Resulta de suma importancia destacar que los reactivos usados en la escala de Likert no son sólo afirmaciones sino que se pueden extender a preguntas u observaciones⁵⁰. La escala de Likert es empleada, en la mayoría de los casos, para medir actitudes con respecto a algún concepto o fenómeno, por ejemplo: el presidente de una nación, en otras ocasiones lo que nos interesa es medir los atributos de un objeto, por ejemplo: la dulzura de un caramelo, la nitidez de una imagen, lo confortable de un asiento, etc.

⁴⁹ Weicrs, R. p. 38

⁵⁰ Hernández, R. p. 26

4.5.5.1 Análisis de reactivos

Aparentemente lo que hemos expuesto hasta este momento resulta sencillo, basta con generar una serie de reactivos y elegir una escala de respuestas, sin embargo, una vez que tenemos el conjunto de reactivos para medir una variable surgen preguntas tales como: ¿Cuáles de esos reactivos serán los que realmente nos permitirán obtener información, válida y confiable con respecto al ente medido?

Likert propone dos métodos para analizar los reactivos y determinar en que medida contribuye cada uno de ellos en la medición total, para de esta forma, hacer la elección correcta, los métodos son:

1. **Método de correlación.**
2. **Método basado en el criterio de consistencia interna.**

4.5.5.1.1 Método de correlación.

En este método se calcula la correlación entre cada reactivo y la suma (o promedio) de todos los reactivos que conforman el instrumento. "Si se obtiene una correlación cercana a cero o muy baja esto indica que el reactivo falla en medir aquello que el resto de los reactivos miden", Likert llama a esta clase de reactivos "reactivos indiferentes", argumentando que, una vez identificados, lo conveniente es retirarlos de la versión final del instrumento pues su aportación para medir a la variable sería mínima por no decir nula. Sin embargo, ya que la puntuación total contiene al reactivo como componente, la correlación entre ambos (puntuación total y reactivo), podría estar inflada artificialmente pues tanto el reactivo como la puntuación total contienen la varianza del reactivo, ante esta observación, se propone una corrección a la fórmula para calcular el coeficiente de correlación, que viene dada por:

$$r_{i(T-i)} = \frac{(r_{iT} \sigma_T - \sigma_i)}{\sqrt{(\sigma_T^2 + \sigma_i^2) - 2r_{iT} \sigma_T \sigma_i}}$$

Donde: $r_{i(T-I)}$: es la correlación entre el reactivo i y el puntaje total de la escala sin el reactivo.

r_{TT} : es la correlación entre el reactivo y el puntaje total.

σ_T : es la desviación estándar de la puntuación total de la escala.

σ_i : es la desviación estándar de la puntuación del reactivo.

En esta fórmula a mayor número de reactivos, cada uno de ellos contribuye en menor peso al puntaje total y por consiguiente menor sesgo es introducido por cada uno de ellos, solo debemos ser precavidos al emplearla cuando su número sea muy reducido.

4.5.5.1.2 Criterio de consistencia interna.

Los pasos que describiremos a continuación se llevan a cabo para analizar los reactivos por medio del criterio de consistencia interna:

- Se genera el conjunto de reactivos.
- Se establecen las categorías de las respuestas, deberán ser las mismas para todos los reactivos.
- Los reactivos son aplicados a una muestra de personas obtenida de la población que será encuestada.
- Las respuestas son codificadas y en base a los resultados obtenidos se separan los reactivos en dos grupos, un grupo está compuesto por los reactivos de las personas que mostraron una actitud más favorable con respecto al objeto o atributo medido, el otro grupo se forma de los reactivos de las personas que expresaron una actitud menos favorable o desagradable con respecto al mismo objeto; usualmente, para hacer la división se toma el 25% de las puntuaciones más altas y el 25% de las puntuaciones más bajas, sin embargo, este porcentaje puede variar.

- Una vez que se han dividido e identificado los 2 grupos digamos g_1 y g_2 , calculamos la media de cada reactivo por grupo obteniendo m_1 y m_2 , respectivamente, restamos una de la otra⁵¹ y así, por último, se obtiene una lista de diferencias. Los reactivos que se elegirán para conformar el instrumento final serán aquellos que tengan las diferencias más altas.

En los Cuadros 4.4.1, 4.4.2, 4.4.3, y 4.4.4 se muestran los resultados obtenidos en los reactivos que se mostraron en los párrafos anteriores, después de haber sido aplicados a una muestra compuesta de 200 personas.

Como bien puede apreciarse, la afirmación número 1 discrimina mejor que la afirmación número 2, ya que su diferencia es mayor, por tanto en caso de tener que elegir por alguno de estos 2 reactivos lo haríamos por el primero. Este método de análisis está sustentado en el hecho de que si un reactivo es adecuado para medir a una variable (en este caso actitudes), será capaz de reflejar las diferencias entre las personas que tienen una actitud más positiva de las que tiene una actitud positiva sobre aquello que estamos midiendo.

Cuadro 4.4.1

Afirmación 1: Es aburrido ver un juego de fútbol.

Respuesta de 50 personas con actitud favorable, sus puntuaciones totales están dentro del 25% superior.

	Total	
Categoría de Respuestas:	Totalmente de acuerdo(5)	0
	De acuerdo.....(4)	0
	Neutral.....(3)	10
	En desacuerdo(2)	20
	Totalmente en desacuerdo.....(1)	20
	Media (X_{s1}):	1.8

⁵¹ La diferencia de estas medias es llamada, por algunos autores, diferencia de valores escalares.

Cuadro 4.4.2

Afirmación 1: Es aburrido ver un juego de fútbol.

Respuesta de 50 personas con actitud desfavorable, sus puntuaciones totales están dentro del 25% inferior.

		Total
Categoría de Respuestas	Totalmente de acuerdo(5)	10
	De acuerdo.....(4)	30
	Neutral.....(3)	5
	En desacuerdo(2)	5
	Totalmente en desacuerdo.....(1)	0
	Media (X_{11}):	3.9

Diferencia de medias para la afirmación 1: $\mu_1 = x_{s1} - x_{11} = 3.9 - 1.8 = 2.1$

Cuadro 4.4.3

Afirmación 2: El fútbol, está próximo a ser el mejor de los espectáculos deportivos.

Respuesta de 50 personas con actitud favorable, sus puntuaciones totales están dentro del 25% superior.

		Total
Categoría de Respuestas	Totalmente de acuerdo(5)	0
	De acuerdo.....(4)	15
	Neutral.....(3)	30
	En desacuerdo(2)	5
	Totalmente en desacuerdo.....(1)	0
	Media (X_{s2}):	3.2

Cuadro 4.4.4

Afirmación 2: El fútbol está próximo a ser el mejor de los espectáculos deportivos.

Respuesta de 50 personas con actitud desfavorable, sus puntuaciones totales están dentro del 25% inferior.

		Total
Categoría de Respuestas	Totalmente de acuerdo(5)	0
	De acuerdo.....(4)	5
	Neutral.....(3)	30
	En desacuerdo(2)	15
	Totalmente en desacuerdo.....(1)	0
Media (X_{12}):		2.8

Diferencia de medias para la afirmación 1: $\mu_2 = x_{s2} - x_{i2} = 3.2 - 2.8 = .4$

El criterio de consistencia interna se ha empleado con mucha frecuencia, no obstante, en la actualidad existe un método alternativo que ha surgido de modificaciones que se le han hecho; el nuevo procedimiento es llamado razón crítica. La razón crítica, o prueba t, evalúa las medias de los grupos en relación con las variancias de los reactivos.

La expresión que nos permite calcular la razón crítica es:

$$t = \frac{(\bar{X}_s - \bar{X}_i)}{\sqrt{\left(\frac{\sigma_s^2}{n_s}\right) + \left(\frac{\sigma_i^2}{n_i}\right)}}$$

Donde: x_i : es la media de la respuesta del reactivo en el grupo i.

σ_i^2 : es la variancia del reactivo en el grupo i

n_i : es el tamaño del grupo i

La razón crítica incluye la distribución de las respuestas por grupo además de la diferencia de medias, motivo por el cual es un indicador más exacto del grado en que los reactivos de actitudes favorables y desfavorables difieren. Si dos reactivos están idénticamente discriminados de acuerdo al criterio de diferencia de medias y los grupos son homogéneos en las respuestas del primer reactivo y mucho menos con respecto al segundo reactivo, la razón crítica es sensitiva al poder de diferenciación del primer reactivo.

Una vez expuestos los dos métodos y explicado en que consisten la pregunta natural que surge es: ¿Qué método elegir?, Likert compara los dos métodos de los que hemos hablado y llega a la siguiente conclusión:

"Hay una relación muy estrecha entre los dos métodos de análisis de reactivos (análisis de correlación y análisis de consistencia interna) pues el coeficiente de correlación ρ (rho) calculado entre ellos es .91"⁵².

Ante el problema de asignar valores a los reactivos de una escala (codificar respuestas) los estudios han demostrado que esencialmente el mismo resultado final se ha obtenido con un arbitrario sentido común de asignación de pesos, que con las más complejas, pero aún arbitrarias, técnicas estadísticas.

En lo que concierne a la confiabilidad del instrumento Likert elige el método Split-Halves (División por Mitades), pues es el que requiere el menor número de cálculos, sin embargo, en estos tiempos el coeficiente α de Cronbach es el estimador de confiabilidad preferido para determinar la confiabilidad de este tipo de escala⁵³.

La presentación de los reactivos que formaran nuestro instrumento se hará al final de la siguiente sección.

⁵² G. Carmines Edward, McIver P. Jhon. "Unidimensional Scaling" Indiana University Sage Publications.

⁵³ Ibid

4.6 PRUEBA PILOTO.

La prueba piloto consiste en aplicar una versión preliminar del instrumento a un grupo determinado de personas (usualmente llamado muestra) con el fin de analizar hasta que punto los elementos que conforman el instrumento funcionan adecuadamente.

Las aplicaciones preliminares del instrumento son convenientes para descubrir ambigüedades y otros problemas potenciales que deberán ser corregidos; a través de este mecanismo se verifica que las instrucciones, los reactivos y las categorías de respuestas sean entendibles, y que el formato del instrumento sea agradable al lector. No existe un número determinado de pruebas que debamos de aplicar antes de poner en marcha nuestro instrumento, quizá sea suficiente con que efectuemos una o bien valga la pena emplear dos, tres o quizá más, lo importante es que el instrumento sea afinado hasta donde sea posible, siempre y cuando recordemos que: "la meta primaria no es obtener un instrumento perfecto, sino estudiar objetivamente una muestra de respondientes extraída de una determinada población"⁵⁴. Dentro de este proceso también se recomienda mostrar el instrumento a personas relacionadas con él o los temas de estudio, pues sus comentarios o sugerencias serán aportaciones valiosas e incuantificables para mejorar el instrumento.

4.6.1 Tamaño de la muestra.

El número de personas a quienes debemos aplicar nuestro instrumento para que el análisis efectuado sobre los datos que obtengamos mediante su aplicación sea satisfactorio, es actualmente un tema de investigación.

Existe una suposición que se hace entorno a la construcción de los instrumentos de medición psicológica estableciendo que, "si pudiéramos medir las diferencias entre los individuos en una escala de intervalos, obtendríamos una

⁵⁴ Weiers, R. p. 38

distribución normal de los puntajes individuales⁵⁵. Si asumimos que los valores de la variable que estamos examinando se distribuyen normalmente, para calcular el tamaño de la muestra podríamos emplear la siguiente fórmula:

$$n = \frac{Z^2 p(q)}{E^2} \quad (4.1)$$

Donde: n: Tamaño de la muestra.

Z: Número de unidades de desviación estándar en la distribución normal que producirá el nivel de confianza deseado.

p: Proporción de la población que posee, la característica en cuestión (se puede estimar o utilizar p=.5).

q: Proporción de la población que no posee la característica en cuestión (q=1-p).

E: Error, o diferencia máxima entre la media de la muestra y la media de la población que estamos dispuestos a aceptar en el nivel de confianza que hemos indicado.

O bien, en caso de conocer el tamaño N de la población, en lugar de la fórmula 4.1 se podría emplear la siguiente ecuación⁵⁶:

$$n = \frac{Z^2 Npq}{E^2 (N - 1) + Z^2 pq} \quad (4.2)$$

Donde: n: Tamaño de la muestra.

⁵⁵ Magnusson, D. p. 68

⁵⁶ La fórmula 4.2 fue empleada en un estudio que realizó la Universidad de Guadalajara, puede consultarse más información al respecto visitando la dirección en Internet que se encuentra al final de estas notas.

- Z:** Número de unidades de desviación estándar en la distribución normal que producirá el nivel de confianza deseado.
- N:** Tamaño de la población.
- p:** Proporción de la población que posee la característica en cuestión (se puede estimar o utilizar $p=.5$).
- q:** Proporción de la población que no posee la característica en cuestión ($q=1-p$).
- E:** Error, o diferencia máxima entre la media de la muestra y la media de la población que estamos dispuestos a aceptar en el nivel de confianza que hemos indicado.

La aplicación de alguna de las dos fórmulas expuestas, para establecer el tamaño de la muestra, requiere que las medidas observadas se distribuyan de manera normal, sin embargo, esta hipótesis no siempre se cumple; razón por la cual se usan fórmulas como⁵⁷:

$$n = 10 * K^v \quad (4.3)$$

Donde: **n:** Tamaño de la muestra.

10: Constante.

K: Promedio de opciones o alternativas de respuesta a los reactivos.

v: Número de reactivos.

En la cual se prescinde de la hipótesis de distribución normal.

⁵⁷ Gorec, Klaus y otros. "Cálculo para evaluar la validez y confiabilidad para evaluar pruebas diagnósticas: Un estudio preliminar", Revista intercontinental de psicología y Educación, 1º de Junio año 1988

Las tres fórmulas arriba descritas no están diseñadas específicamente para calcular muestras que conduzcan a determinar la validez y la confiabilidad del instrumento, empero, son las que más se acercan a las necesidades de los investigadores, y por tanto, dichas fórmulas son consideradas como un primer recurso para establecer el tamaño de la muestra que se extraerá de la población.

Algunos investigadores se han dado a la tarea de desarrollar fórmulas para determinar cual es el *tamaño mínimo de la muestra* de personas que deben ser encuestadas para que los coeficientes de validez y confiabilidad calculados a partir de los datos recabados sean, invariablemente, indicadores de que el instrumento construido cumple con todos los requerimientos esperados, resultados de sus estudios han arrojado expresiones como:

$$n = 10 * (V * m) \quad (4.4)$$

- Donde:
- n:** Tamaño de la muestra.
 - V:** Número de reactivos incluidos en el instrumento.
 - m:** Número promedio de alternativas o posibles respuestas.
 - 10:** Una constante obtenida de las limitaciones implícitas de la prueba de significancia chi cuadrada.

Esta última fórmula, está basada esencialmente en dos elementos que son afines a todos los instrumentos de medición: el número de reactivos y el promedio de opciones de respuesta. Es de suma importancia destacar que esta ecuación ha demostrado su eficacia tomando como ejemplo una prueba diagnóstica de alcoholismo. La prueba consistió en calcular el tamaño de la muestra para este test para lo cual se usaron las fórmulas 4.1 y 4.4, después de hacer ciertos ajustes al instrumento, se aplicó, se recabaron los resultados y se calcularon coeficientes de validez (validez concurrente) y coeficientes de confiabilidad (métodos KR20 y Spearman-Brown); los resultados fueron altamente satisfactorios.

La documentación que actualmente existe para el tema que hemos discutido en esta sección no es tan vasta y extensa como para los otros temas de los cuales se ha hablado a lo largo de la presente obra, sin embargo, la mayoría los estudios de mercado coinciden en que, entre menor sea el tamaño de la muestra a examinar, menores serán los gastos empleados para realizar una encuesta, este es uno de los motivos principales por los cuales bien valdría la pena realizar más trabajos de investigación al respecto.

4.7 EL INSTRUMENTO

Ahora que se han analizado los pasos para construir un instrumento de medición presentaremos los reactivos que nos permitirán medir las variables que hemos definido, no sin antes tomar en cuenta ciertas consideraciones.

4.7.1 Recomendaciones para la redacción de los reactivos.

Las encuestas se pueden manipular, esto es una afirmación, la pregunta es ¿Cómo se puede alterar una encuesta?. La respuesta a esta interrogante nos la da el encuestador Michael Marsolini: "La forma de preguntar es muy importante, si las preguntas (reactivos) se hacen de cierta forma se obtienen respuestas muy diferentes, cada coma, cada cambio de palabra es importante. Hay una historia que se relata en el medio de la investigación de mercados sobre un sacerdote reñido por sus superiores cuando preguntó si podía fumar mientras oraba, un amigo le dijo que había formulado mal la pregunta, que ésta debió ser: ¿Puedo orar mientras fumo? ". Ante tal observación se ha estipulado seguir algunas reglas prácticas para la formulación de los reactivos, algunas de ellas son⁵⁸:

- Usar palabras sencillas, el reactivo debe ser de un vocabulario familiar para el sujeto que conteste el cuestionario.

⁵⁸ Ander-Egg, Ezequiel "Técnicas de investigación social" Ed. El Ateneo. 21 Edición. México, 1989.

- El reactivo debe ser lo más preciso posible, de preferencia no debe exceder de 20 palabras.
- Evitar los reactivos ambiguos, es decir, que puedan interpretarse de diferentes maneras.
- Cada reactivo debe poseer una idea.
- Evitar la tendencia (direccionalidad) en los reactivos.
- Evitar usar las palabras universales: "siempre, ninguno, nunca", estos solo se usan para definir los extremos del continuo, o aquellas como "solo, aquellas, casi, a veces", porque introducen ambigüedad, pues significan diferentes cosas para diferentes personas.
- Procurar cubrir el dominio o dimensión de la variable en todo el rango definido.
- Un reactivo nunca debe empezar con una negación.

4.7.2 Lista de reactivos

De acuerdo con las variables que se establecieron en secciones anteriores y tomando como punto de partida un conjunto de preguntas generado por personal especializado de la DGSCA (pedagogas, trabajadoras sociales, psicólogas, etc.) así como su previa revisión; se generó una lista de reactivos la cual se presenta a continuación (los reactivos están clasificados de acuerdo a la variable a la que pertenecen.:

• Dominio de conocimientos sobre la materia.

1. El instructor explicaba claramente los temas.
2. La forma en que el instructor abordaba los temas facilitaba la asimilación de los conocimientos.

• **Estrategias de exposición y conducción de la clase**

3. Las prácticas realizadas en el salón de clases facilitaban el aprendizaje de los temas tratados
4. La secuencia de los temas permitía una mejor comprensión de los mismos

• **Relación del profesor con los alumnos.**

5. El instructor mostraba buena disposición para atender las dudas de los alumnos
6. El instructor propiciaba el planteamiento de dudas y aportaciones acerca de los temas vistos en clase.

• **Cumplimiento del programa oficial del curso y logro de los objetivos propuestos.**

7. Los temas vistos en clase permitieron el cumplimiento de los objetivos.

• **Procedimiento para evaluar el aprendizaje de los alumnos.**

8. Las evaluaciones del instructor iban de acuerdo a los temas vistos en clase.

• **Empleo de recursos y medios didácticos.**

9. Los medios y recursos didácticos ilustraban adecuadamente los temas.

• **Infraestructura.**

10. El funcionamiento del equipo de cómputo permitía el buen desarrollo del curso.
11. Las condiciones de las instalaciones eran adecuadas para el curso.

4.7.3 Presentación del Instrumento

El aspecto final del instrumento: instrucciones, reactivos, escala de respuestas y formato se presenta en el Cuadro 4.5 .

4.7.4 Observaciones

En lo que concierne a las categorías de respuesta, se ha adoptado lo siguiente:

(5) Siempre o Casi Siempre.

(4) Muchas veces.

(3) Regularmente.

(2) Pocas veces.

(1) Nunca o casi nunca.

Como puede notarse todos los reactivos son afirmaciones positivas con respecto a alguna característica por evaluar, por tanto, la codificación de las respuestas se ha hecho asignando los pesos más altos a las respuestas más favorables, resulta conveniente mencionar que estamos trabajando básicamente con una escala tipo Likert. El valor que le corresponde a cada categoría de respuestas se encuentra dentro de los paréntesis, a la hora de aplicar el cuestionario deberán omitirse.

CUADRO 4.5
FORMATO DEL CUESTIONARIO

Instrucciones

Con el fin de mejorar nuestros servicios le solicitamos que nos ayude a contestar, de manera objetiva, este pequeño cuestionario. Las opiniones de todos los encuestados serán resumidas en informes finales, sus respuestas serán confidenciales y anónimas. Por favor exprese su opinión marcando con una (X) su respuesta.

Reactivo	Nunca o casi nunca	Pocas veces	Regularmente	Muchas veces	Siempre o casi siempre
1. El instructor explicaba claramente los temas.	()	()	()	()	()
2. La forma en que el instructor abordaba los temas facilitaba la asimilación de los conocimientos.	()	()	()	()	()
3. Las prácticas realizadas en el salón de clases facilitaban el aprendizaje de los temas tratados.	()	()	()	()	()
4. La secuencia de los temas permitía una mejor comprensión de los mismos.	()	()	()	()	()
5. El instructor mostraba buena disposición para atender las dudas de los alumnos.	()	()	()	()	()
6. El instructor propiciaba el planteamiento de dudas y aportaciones acerca de los temas vistos en clase.	()	()	()	()	()
7. Los temas vistos en clase permitieron el cumplimiento de los objetivos.	()	()	()	()	()
8. Las evaluaciones del instructor iban de acuerdo a los temas vistos en clase.	()	()	()	()	()
9. Los medios y recursos didácticos ilustraban adecuadamente los temas.	()	()	()	()	()
10. El funcionamiento del equipo de cómputo permitía el buen desarrollo del curso.	()	()	()	()	()
11. Las condiciones de las instalaciones eran adecuadas para el curso.	()	()	()	()	()

Comentarios

GRACIAS POR SU COLABORACIÓN

Las categorías de respuestas están equilibradas, es decir, hay igual número de categorías favorables como desfavorables, con lo cual se pretende evitar la tendencia direccional, además, como el número de alternativas es impar, se ha establecido dejar una opción neutral, con el fin de que un alumno que realmente desee expresar esa posición lo pueda hacer, en caso de que los resultados presenten una alta tendencia hacia esta alternativa, lo recomendable sería, antes de hacer alguna otra cosa, quitarla del cuestionario y aplicarlo nuevamente. Las categorías de respuestas tienen descripciones verbales pretendiendo que sean lo más equidistantes posibles desde el punto de vista psicológico (claras y precisas).

Con el fin de recuperar la mayor cantidad de información posible, en la parte final del cuestionario se ha agregado una sección de comentarios, para que el alumno pueda hacer observaciones que contribuyan a la mejora de nuestro estudio.

El instrumento es de tipo directo estructurado, los encuestados conocen el propósito de la encuesta y el proceso de aplicación será el mismo en todos los casos. Las personas que serán encuestadas serán los alumnos que asisten a los cursos, es decir, los estudiantes; que como han mostrado las investigaciones efectuadas, son la mejor fuente de información a la cual se puede recurrir para evaluar un curso.

4.8 ANÁLISIS DEL INSTRUMENTO.

El número de cuestionarios aplicados fue 103, de los cuales para el análisis final se descartaron 6 debido a la omisión en algunas respuestas por parte de los estudiantes, quedando así un total de 97. Dado que la escala del cuestionario es de tipo Likert se pudieron efectuar los estudios que se describen a continuación:

4.8.1 Análisis de Reactivos

- Para este análisis se emplearon el criterio de consistencia interna y el método de correlación descritos en secciones anteriores. (los resultados se encuentran en las columnas del Cuadro 4.6)
- En lo que concierne al criterio de consistencia interna se aplicó el método tradicional que consiste en separar el total de cuestionarios en dos grupos considerando la puntuación total, la división se hizo tomando en cuenta el 25% de las puntuaciones más altas y el 25% de las puntuaciones más bajas. Se calculó la media de todos los reactivos de cada grupo y se tomó la diferencia (Columna 3). También se calculó la razón crítica (Columna 5) para cada reactivo, esto nos servirá para poder comparar ambos procedimientos ya que el fin que persiguen es el mismo: identificar qué reactivos son los que discriminan mejor las respuestas favorables de las desfavorables.
- Los métodos de correlación para el análisis de reactivos también se muestran en el Cuadro 4.6, primero se obtuvo la correlación del i -ésimo ítem con el puntaje total del cuestionario (Columna 7) y después se calculó la correlación entre el i -ésimo ítem y el puntaje total de cuestionario después de eliminar el i -ésimo ítem (Columna 9).
- Para el análisis de reactivos mediante los métodos Poder de Discriminación y Razón Crítica ordenamos los datos de manera descendente y asignamos rangos, los resultados nos muestran que las columnas 4 y 6 no coinciden para ningún reactivo, siendo el reactivo R10 el que presenta mayor contraste pues en un método ocupó el primer lugar y en otro el séptimo, el coeficiente ρ de Spearman calculado para estos rangos es de tan solo de 0.5727.
- También se asignaron rangos para los resultados obtenidos del cálculo de correlaciones, los reactivos R10 y R11 son los que muestran menor correlación con el puntaje total, el coeficiente ρ de Spearman calculado para estos rangos es de 0.9909 con lo cual podemos concluir la congruencia de estos métodos de análisis.

CUADRO 4.6
ANÁLISIS DE REACTIVOS

1	2	3	4	5	6	7	8	9	10
Ítem	Media	Poder Discriminación	Rango	Razón Crítica	Rango	Correlación Ítem-Total	Rango	Correlación Ítem - Total Sin el Ítem	Rango
R1	4.61	1.33	3	10.47	1	0.8458	2	0.8099	2
R2	4.52	1.38	2	8.94	3	0.7541	9	0.6921	9
R3	4.65	1.13	10	6.26	11	0.8227	4	0.7817	4
R4	4.52	1.33	4	9.51	2	0.8366	3	0.7973	3
R5	4.69	1.21	7	6.84	8	0.8478	1	0.8117	1
R6	4.58	1.29	5	6.77	9	0.7896	7	0.7353	8
R7	4.53	1.21	8	7.76	6	0.8005	5	0.7550	5
R8	4.68	0.92	11	6.37	10	0.7910	6	0.7522	6
R9	4.55	1.21	9	8.39	4	0.7885	8	0.7396	7
R10	4.29	1.58	1	7.20	7	0.6991	10	0.6039	10
R11	4.53	1.25	6	7.88	5	0.6424	11	0.5633	11

Antes de emitir juicio alguno bien valdría la pena hacer mención de dos cosas:

- La decisión de un método para evaluar los reactivos no debe tomarse a la ligera.
- Revisaremos los resultados obtenidos en el análisis de confiabilidad.

4.8.2 Análisis de Confiabilidad

En lo que concierne al análisis de confiabilidad se obtuvo la matriz de correlaciones así como el coeficiente alfa de Cronbach.

La matriz de correlaciones se presenta en el Cuadro 4.7

Examinando la matriz de correlaciones claramente se revela la debilidad de los reactivos R10 y R11. Las correlaciones entre éstos reactivos y los restantes son las más bajas, de hecho están por debajo del promedio ($p = .5724$).

Los cálculos del alpha de Cronbach y la varianza total del instrumento, ambos, sin considerar el i -ésimo ítem, se muestran en el Cuadro 4.8.

CUADRO 4.7
MATRIZ DE CORRELACIONES

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
R1	1										
R2	0.7228	1									
R3	0.7038	0.6533	1								
R4	0.6404	0.5733	0.6762	1							
R5	0.7116	0.6673	0.7316	0.7360	1						
R6	0.6927	0.5207	0.5981	0.6729	0.6076	1					
R7	0.6410	0.5359	0.6090	0.6795	0.6023	0.6564	1				
R8	0.6516	0.5496	0.5979	0.7644	0.7083	0.5984	0.6556	1			
R9	0.6231	0.5047	0.6565	0.6434	0.6698	0.6005	0.6016	0.6855	1		
R10	0.4813	0.4158	0.4259	0.4586	0.4993	0.4924	0.5280	0.3962	0.4461	1	
R11	0.4634	0.3639	0.4978	0.4420	0.4300	0.3628	0.4204	0.3550	0.4312	0.6464	1

CUADRO 4.8

ANÁLISIS DE CONFIABILIDAD

Item i	Varianza de la escala Sin el Item i	Alpha de Cronbach Sin el Item i
R1	31.3149	0.9225
R2	31.3241	0.9274
R3	31.4394	0.9236
R4	31.1783	0.9228
R5	31.1856	0.9223
R6	31.0213	0.9254
R7	31.6179	0.9247
R8	32.6244	0.9258
R9	31.6007	0.9253
R10	30.4517	0.9349
R11	32.5554	0.9329

Número de Casos : 97

Alpha = .9324

El alpha total del instrumento, considerando los 11 reactivos y los 97 casos es .9324, que es menor al alpha calculada si elimináramos los reactivos R10 y R11 de acuerdo con el Cuadro 4.8.

4.9 IMPLEMENTACION DEL CUESTIONARIO.

Para la implementación del cuestionario deberá seguirse el siguiente procedimiento :

- Presentar el cuestionario a la Subdirección de la DGSCA.

- **Si es aceptado:**
 - **Será llevado al Departamento de Desarrollo Curricular para su reproducción.**
 - **La sustitución del nuevo cuestionario por el anterior será de forma paulatina en lo que se modifica el sistema.**
 - **También sufrirán modificaciones los reportes a profesores y los informes de cursos que se entregan a la Dirección.**
- **Si no es aceptado**
 - **El estudio habrá terminado en la etapa de análisis.**

CONCLUSIONES

En la actualidad las encuestas se han convertido en un elemento más para tomar decisiones importantes, resulta ser muy familiar para nosotros el que se nos someta a una serie de preguntas con respecto a algún tema vigente o de gran interés ya sea político, deportivo, cultural, etc., como muestra tenemos las preguntas que lanzan los noticieros nacionales en sus turnos en vivo, los plebiscitos que últimamente están de moda en el Distrito Federal e incluso el Censo de Población y Vivienda que está a cargo del INEGI.

La necesidad de construir instrumentos de medición que demuestren su confiabilidad y validez es evidente dada la trascendencia que tienen las resoluciones que se toman a partir de los datos obtenidos por dichos instrumentos.

A lo largo de esta obra se describe la metodología para construir un instrumento de medición y más aún se ha construido uno de ellos basándose en la teoría existente cuyo objetivo es evaluar los cursos de cómputo que imparte la Dirección General de Servicios de Cómputo Académico (DGSCA) de la UNAM culminando con un análisis de confiabilidad.

Los resultados del análisis son satisfactorios y por tanto se concluye que los datos obtenidos mediante la aplicación del instrumento son confiables.

Es importante mencionar que al comenzar el presente trabajo nunca imaginé que la construcción de un instrumento, en particular, un cuestionario, llevara de fondo una teoría matemática tan interesante, basada, principalmente, en análisis estadístico. La experiencia obtenida ha sido muy enriquecedora y abre una alternativa más para las funciones que un actuario puede desempeñar en el campo laboral.

GLOSARIO

Análisis de ítems. Proceso consistente en evaluar ítems de prueba aislados con respecto a ciertas características. Suele implicar la determinación de dificultad y el poder de discriminación del ítem, y muchas veces, su correlación con algún criterio externo.

Análisis Factorial. Cualquiera de los varios métodos que se utilizan para analizar las inter-correlaciones existentes entre una serie de variables. El análisis factorial intenta explicar las interrogantes existentes en términos de algunos "factores" subyacentes, cuyo número es de preferencia menor que el de las variables originales, y revela qué proporción de la variación observada en cada una de las medidas originales se debe a o se asocia con cada uno de los factores hipotéticos. Cabe agregar que el análisis factorial ha contribuido a una mejor comprensión de la organización de los componentes de la inteligencia, entre otras cosas.

Aprendizaje. Se define comúnmente como la modificación de la conducta que se produce como resultado de una experiencia, entendiéndose a ésta como un proceso de interacción entre el individuo y su ambiente, que se traduce en conocimientos, actitudes y destrezas que el individuo adquiere.

Aptitud. Combinación de destrezas y otras características, ya sean innatas o adquiridas, indicadoras de la capacidad de un individuo para aprender o adquirir competencia en determinada área, siempre y cuando se le proporcione la educación o el entrenamiento necesarios.

Algunos definirían a "aptitud" en un sentido mas integral. Así "aptitud musical" se referiría no sólo a la combinación de características físicas y mentales, sino también a factores motivacionales, intereses y posiblemente otras características, susceptibles de facilitar el aprendizaje en el área musical.

Autoevaluación. Apreciación de los resultados del aprendizaje o de cualquier otra actividad hecha por el sujeto mismo que aprende o actúa, apreciación que se expresa mediante un juicio de valor.

Batería. Grupo de varias pruebas estandarizadas con una misma muestra de la población de modo que son comparables los resultados obtenidos entre ellas. A veces se aplica en un sentido general a cualquier grupo de pruebas administradas juntas, aunque éstas no hayan sido estandarizadas con los mismos sujetos. Las baterías de pruebas más comunes son las de rendimiento escolar, que incluyen subtesis en las diferentes áreas de aprendizaje.

Calificar. Expresar en términos cualitativos el resultado de una medición o una apreciación, como en el caso de los datos que aporta un cuestionario, una escala estimativa, etc., cuando se resumen en un término que puede ser una cantidad o una expresión calificativa, como bueno, muy bueno, etc.

Consistencia. Tratándose de las pruebas denominadas objetivas, calidad que asegura que sus resultados no variarán a través de sucesivas aplicaciones de un instrumento de medición.

Coefficiente de confiabilidad. Coeficiente de correlación entre dos formas de una prueba, entre las puntuaciones obtenidas en dos administraciones de la misma prueba, o entre las dos mitades de una prueba, debidamente corregida.

Coefficiente de correlación. Medida del grado de relación o de "parentesco" entre dos series de medidas provenientes del mismo grupo de individuos. El coeficiente de correlación más comúnmente usado en la elaboración de pruebas y la investigación en el área educativa es el que se conoce bajo el nombre de coeficiente de Pearson o de Producto-Momento. Los coeficientes de correlación fluctúan entre 0 que indica una falta total de relación y 1 así como -1 indicando una correspondencia perfectamente positiva o perfectamente negativa, respectivamente.

Correlación. Relación o "parentesco" entre dos series de puntuaciones o medidas; tendencia que muestra una puntuación a variar en concordancia con la otra. La existencia de una fuerte relación entre dos variables no necesariamente indica que una de ellas tenga influencia causal sobre la otra.

Desviación estándar. Una medida de la variabilidad o dispersión de una distribución de puntuaciones. A mayor agrupación de las puntuaciones en torno a la media, más pequeña es la desviación estándar.

Distribución (Distribución de frecuencias). Tabulación de las puntuaciones (u otros atributos) de un grupo de individuos, en que se muestra el número (frecuencia) de cada puntuación, o de las incluidas dentro de cada intervalo.

Escala. Conjunto de puntos determinados según cierta ley y dibujados sobre una línea recta o curva que recibe la denominación de soporte de la escala. La escala se construye marcando pequeños trazos normales a su "soporte" y escribiendo frente a ellos una creciente sucesión de números, que representan los valores sucesivos de cierta variable (un metro, un termómetro, etc., son ejemplos de escalas).

Cuando en este texto se habla de escala, se alude a una serie numérica que muestra los diferentes valores que puede tener una variable; por ejemplo, el rendimiento pedagógico de un alumno, la fatiga, el tiempo invertido, etc.

Factor. En la medición mental, rasgo, capacidad o componente de capacidad hipotético que subyace a la ejecución en dos o más pruebas y también influye sobre ésta, haciendo por lo tanto, que se correlacionen las puntuaciones alcanzadas en dichas pruebas. El término factor se refiere exclusivamente a una variable teórica, derivada mediante un proceso de análisis factorial, llevado a cabo con la ayuda de una tabla de Inter-correlaciones entre las pruebas. Sin embargo, se utiliza igualmente para designar la interpretación psicológica dada a la variable, como la destreza verbal, la habilidad numérica, etc.

Inventario. Cuestionario o lista de verificación, generalmente con forma de auto-reporte, destinado a obtener información no intelectual acerca de un individuo. Sin ser pruebas, en el sentido usual de la palabra, los inventarios se refieren en la mayoría de los casos a rasgos de la personalidad, intereses, actitudes, problemas, motivación, etc.

Ítem. Pregunta o ejercicio aislado en una prueba.

Media Aritmética. Tipo de promedio generalmente llamado media. Se calcula dividiendo la suma de una serie de puntuaciones por el número de puntuaciones.

Muestra Representativa. Muestra que corresponde o iguala a la población de la que fue sacada con respecto a características importantes para los propósitos de la investigación que se realiza.

Percentil (P). El punto (puntuación) en una distribución en el cual o debajo del cual se encuentra el porcentaje de casos indicado por el percentil. Así, una puntuación que coincide con el 35º percentil (P_{35}) se considera igual o superior a la del 35% de las personas del grupo, de modo que el 65% de las ejecuciones superan dicha puntuación.

Proceso de Enseñanza Aprendizaje. Conjunto de las fases sucesivas del fenómeno en el que intervienen como elementos un alumno, un contenido, (conocimiento, actitud, destreza por aprender) y, un guía, que en los sistemas tradicionales está representado por el profesor.

La enseñanza y el aprendizaje contemplados a la luz de la didáctica moderna, constituyen un proceso intencionado y sistemático que se inicia con el planteamiento de un propósito concreto y definido, y concluye con la ejecución de una nueva conducta esperada del alumno, hecho que a su vez es tomado en cuenta para modificar el proceso si resulta necesario.

Rango. Para algún grupo específico, la diferencia entre la más alta y la más baja de las puntuaciones obtenidas en una prueba; por lo tanto una medida muy burda de

expansión o variabilidad ya que solamente se basa en dos puntuaciones extremas. El termino rango se utiliza igualmente en lo que se refiere a la posible dispersión de la medición que permite una prueba, que en la mayoría de los casos, corresponde al número de ítems de la prueba.

Reactivo. Instancia o aspecto elemental que se consigna en una prueba y que está diseñado para provocar una respuesta de parte del alumno.

Test. Conjunto de métodos que permiten valorar o medir una o varias características de un individuo, un grupo, un producto o una máquina.

Tratamiento estadístico. Conjunto de operaciones matemáticas que permite describir y calificar las relaciones y los valores que adoptan las variables o atributos que se estudian en un conjunto de individuos.

APÉNDICE I

MEDIDAS DE CORRELACIÓN.

Una medida de correlación es un índice que expresa el grado de asociación entre 2 variables X y Y y que debe poseer las siguientes propiedades:

- a) El valor de la medida de correlación oscila entre -1 y 1.
- b) Si los valores más grandes de X tienden a estar apareados con los valores más grandes de Y y si los valores más pequeños de X tienden a estar apareados con los valores más pequeños de Y, entonces el valor de la medida de correlación deberá ser positivo y cercano al valor 1; cuando lo anterior ocurre se dice que existe una correlación positiva entre las variables X y Y, de tal forma que si la medida de correlación calculada entre éstas da el valor 1 se diría que existe una correlación perfecta positiva entre ambas variables.
- c) Si los valores más pequeños de X tienden a estar apareados con los valores más grandes de Y y si los valores más grandes de X tienden a estar apareados con los valores más pequeños de Y, entonces el valor de la medida de correlación será negativo y cercano a -1, cuando esto último ocurre se dice que existe una correlación negativa entre X y Y, de tal forma que si la medida de correlación calculada entre éstas variables da el valor -1 se diría que hay una correlación perfecta negativa entre ambas.
- d) Si los valores de la variable X se aparean aleatoriamente con los valores de la variable Y, el valor de la medida de correlación será cercano a cero.

Un problema de correlación se presenta cuando queremos indagar si existe alguna relación entre un par de variables de interés. Por ejemplo, ¿existe alguna relación entre el uso del tabaco y las afecciones cardiacas?, ¿entre la capacidad para aprender música y la aptitud científica?, ¿entre la cantidad de trabajo realizado por un obrero y la de material perdido?.

Para efectos de nuestro estudio revisaremos dos medidas de correlación:

- El coeficiente de correlación producto-momento (Υ).
- La rho de Spearman (ρ).

Los dos coeficientes de correlación arriba citados nos permiten medir la relación lineal entre dos variables, ya que dada la naturaleza de nuestra investigación, esta clase de medidas es lo que necesitamos para validar los coeficientes que calcularemos. Los resultados de diversos estudios han mostrado que, para muestras grandes ($n \geq 30$, donde n es el tamaño de la muestra), el empleo de la Υ de Pearson genera muy buenos resultados cuando de relación lineal se trata, en complemento, cuando el tamaño de la muestra es pequeño ($n < 30$), podemos recurrir a la rho (ρ) de Spearman. Cada una de estos coeficientes posee ciertas características que nos hace recurrir a uno u otro de acuerdo a nuestras necesidades.

El coeficiente de correlación producto momento (Υ).

El coeficiente de correlación producto-momento, Υ , fue desarrollado por Karl Pearson, motivo por el cual este coeficiente también es conocido como el coeficiente Υ de Pearson. Se ha demostrado que el coeficiente Υ cumple con las 4 propiedades que enunciamos en la sección anterior para ser considerado una medida de correlación.

El coeficiente Υ se calcula a partir de un conjunto de datos que provienen de una muestra aleatoria continua bivariada de tamaño n y que está agrupada por parejas ordenadas de la forma (x_i, y_i) desde $i=1$ hasta n , donde x_i representa la i -ésima observación de la variable X y y_i la i -ésima observación de la variable Y .

La expresión que nos permite calcular el coeficiente Υ de Pearson es:

$$r = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) \left(\sum_{i=1}^n (y_i - \bar{y})^2\right)}} \quad (1)$$

Donde: **Cov(x,y):** Es la covarianza entre las variables X y Y.

$\sigma_x \sigma_y$: Es el producto de la desviación estándar de la variable X por la desviación estándar de la variable Y respectivamente.

Sin embargo, para facilitar el cálculo de Y, emplearemos la expresión (2), que es simplemente una versión equivalente para la expresión (1) y viene dada por:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n (x_i)^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n (y_i)^2 - \left(\sum_{i=1}^n y_i\right)^2}} \quad (2)$$

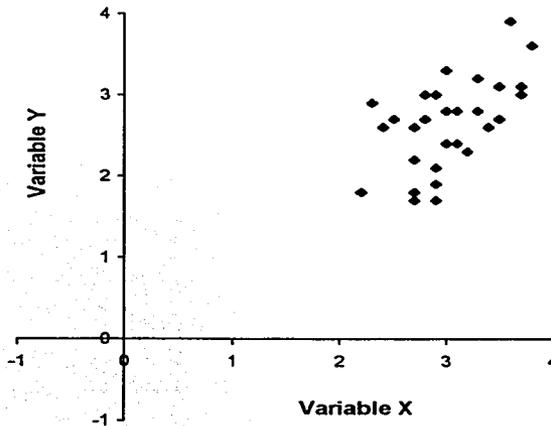
El coeficiente Y nos indica si existe o no una relación **lineal** entre dos variables X y Y, por tal motivo, antes de proceder a calcularlo, debemos asegurarnos de que pueda existir alguna relación de tipo lineal entre ambas variables. Para darnos una idea de que tan cercanos están nuestros datos de mantener una relación lineal podemos proceder a graficarlos. Para efectos de ejemplo supongamos que tenemos las calificaciones de un grupo de 30 estudiantes al final de la preparatoria (variable X) y sus calificaciones al final del primer año universitario (variable Y). El conjunto de datos, parejas de la forma (x,y), se presenta en la Tabla I.

La representación en el plano cartesiano del conjunto de datos de la tabla 1 recibe el nombre de "Gráfica de Dispersión" (Gráfica 1), y podemos visualizarlo como una simple "nube de puntos".

TABLA I					
(X,Y)	(X,Y)	(X,Y)	(X,Y)	(X,Y)	(X,Y)
(3.0,2.4)	(2.9,3.0)	(2.9,1.9)	(2.8,2.7)	(3.1,2.8)	(2.7,1.7)
(2.4,2.6)	(3.5,3.1)	(2.7,2.2)	(3.1,2.4)	(3.3,3.2)	(2.9,1.7)
(3.7,3.0)	(3.0,2.8)	(3.7,3.1)	(2.8,3.0)	(2.7,1.8)	(3.2,2.3)
(3.6,3.9)	(2.3,2.2)	(2.7,2.6)	(3.0,3.3)	(3.5,2.7)	(3.4,2.6)
(3.8,3.6)	(3.0,2.9)	(3.3,2.8)	(2.2,1.8)	(2.9,2.1)	(2.5,2.7)

El Diagrama de Dispersión nos da una idea aproximada de que tan cercanos están nuestros puntos en el plano cartesiano de ajustarse a una línea recta. La observación de este diagrama de dispersión muestra que existe una tendencia de los valores pequeños de X a estar asociados con los valores pequeños de Y, y que los valores grandes de X tienden a estar asociados con los valores de grandes de Y.

Gráfica de Dispersión de los datos de la tabla 1



Una vez graficados los datos y después de identificar de manera intuitiva que éstos se pueden ajustar a una línea recta, procedemos a calcular el coeficiente de correlación de Pearson.

Sustituyendo los valores antes calculados en la expresión (2) obtenemos:

$$r = \frac{30(242.52) - 90.6(78.9)}{\sqrt{30(278.7) - 8208.36} \sqrt{30(216.4) - 6225.21}}$$

Por lo tanto:

$$Y=.63$$

Observaciones

- a) Debemos ser precavidos al momento de interpretar el valor de Y ya que éste simplemente nos da información matemática pura acerca de la relación lineal entre dos variables, no nos proporciona implicaciones acerca de *causa y efecto*⁵⁹. Un analista que obtenga alto grado de correlación entre la natalidad de la India y el producto nacional bruto de los Estados Unidos, debe pensarlo detenidamente antes de llegar a la conclusión de que una cosa es *causa* de la otra. Ante esta observación resulta altamente recomendable utilizar la información que nos proporciona el valor de Y como un elemento más en el contexto de nuestro estudio.
- b) El valor de Y no debe interpretarse como una proporción, un valor de .5 no es la mitad entre 0 y 1 en términos de la exactitud de predicción que puede ser perfecta, solo podemos decir que los valores más grandes de Y (ya sean cercanos a 1 o a -1) dan predicciones más exactas que los valores pequeños o cercanos a 0.

Tomando en cuenta las observaciones antes citadas, el valor $Y=.63$ que obtuvimos en nuestro ejemplo deberá ser analizado dentro del contexto de nuestro

⁵⁹ Paul, Hoel "Estadística Elemental" Ed. Cia Editorial Continental S.A. de C.V., México, 1975, 4ª Edición.

problema antes de emitir un juicio o tomar alguna otra decisión con base a este resultado.

La ρ (rho) de Spearman.

La ρ de Spearman es un coeficiente de correlación de rango, este coeficiente es simplemente el coeficiente de correlación de Pearson calculado para los rangos de dos variables y no para los valores numéricos de estas (siempre y cuando no exista empate alguno entre dichos rangos). Sustituyendo los rangos de las variables X y Y en la expresión (1) obtenemos que la ρ de Spearman adquiere la siguiente forma:

$$\rho = \frac{\sum_{i=1}^n \left[R(x_i) - \frac{\sum_{i=1}^n R(x_i)}{n} \right] \left[R(y_i) - \frac{\sum_{i=1}^n R(y_i)}{n} \right]}{\sqrt{\left(\sum_{i=1}^n \left[R(x_i) - \frac{\sum_{i=1}^n R(x_i)}{n} \right]^2 \right) \left(\sum_{i=1}^n \left[R(y_i) - \frac{\sum_{i=1}^n R(y_i)}{n} \right]^2 \right)}} \quad (3)$$

Donde: $R(x_i)$: Es el i-ésimo rango de la variable X.

$R(y_i)$: Es el i-ésimo rango de la variable Y.

n: Es el número de parejas de la forma $[R(x_i), R(y_i)]$

Realizando simplificaciones puramente algebraicas puede mostrarse que el coeficiente ρ adquiere la siguiente forma:

$$\rho = 1 - 6 \left(\frac{\sum_{i=1}^n (R(x_i) - R(y_i))^2}{n(n^2 - 1)} \right) \quad (4)$$

o bien:

$$\rho = 1 - 6 \left(\frac{\sum_{i=1}^n (d_i)^2}{n(n^2 - 1)} \right) \quad (5)$$

Donde: $d_i = R(x_i) - R(y_i)$: es la i -ésima diferencia entre el rango del i -ésimo dato de la variable X y el i ésimo dato de la variable Y .

n : Es el número de parejas de la forma $[R(x_i), R(y_i)]$.

Ya que mediante la fórmula para obtener el valor de γ se obtiene el valor de ρ , tenemos que éste último también hereda las propiedades necesarias para ser una medida de correlación (la manera de demostrarlo es análoga a la forma en la que se hace para γ) y además su interpretación es la misma.

El procedimiento mediante el cual se obtiene los valores de los rangos tanto de X como de Y [$R(x)$ y $R(y)$, respectivamente], a partir de los valores de X y Y se ilustra a continuación:

Supongamos que tenemos el siguiente conjunto de datos:

Cuadro A

x_i	y_i
7	4
5	7
8	9
9	8

Al convertirlos en una escala ordinal obtendríamos alguno de los siguientes resultados:

Cuadro B

$R(x_i)$	$R(y_i)$
2	1
1	2
3	4
4	3

Cuadro C

$R(x_i)$	$R(y_i)$
3	4
4	3
2	1
1	2

Lo primero que hacemos es ubicar el valor más pequeño de X (Cuadro A), en este caso es el 5, a este valor le asignaremos el rango 1 (Cuadro B), después buscamos el segundo valor más pequeño de X (Cuadro A), en este caso es el 7, que le corresponde el rango 2 (Cuadro B) y así sucesivamente hasta terminar con todos los valores de X , y posteriormente continuamos de manera similar con Y (el resultado final se observa en el Cuadro B). De igual forma, pudimos haber asignado los rangos en orden inverso, es decir, primero identificamos el valor más grande de X (Cuadro A) en este caso resulta ser el 9 y a éste se le asigna el rango 1 (Cuadro C), después buscamos el segundo valor más grande de X (Cuadro A) que es el 8 y le asignamos el rango 2 (Cuadro C) etc., y proceder análogamente con los valores de Y , al final del proceso obtendríamos el Cuadro C. Cualquiera de los dos métodos antes descritos

es correcto, siempre y cuando apliquemos el mismo criterio de asignación a las 2 variables.

En caso de que existan empates a la hora de asignar rangos a las variables, en lugar de emplear la expresión (5) utilizaríamos la siguiente fórmula:

$$\rho = \frac{\sum_{i=1}^n R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^n (R(x_i))^2 - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^n (R(y_i))^2 - n\left(\frac{n+1}{2}\right)^2\right)}} \quad (6)$$

y a los rangos empatados les asignaríamos el promedio del rango que les corresponde, para ilustrar lo anterior supondremos que obtuvimos las siguientes observaciones:

CUADRO A1

x_i	5	3	2	3	7	5	10
y_i	1.5	6	4.5	3	1.5	4.5	7

Después de asignar los rangos obtendríamos:

CUADRO B1

$R(x_i)$	4.5	2.5	1	2.5	6	4.5	7
$R(y_i)$	1.5	6	4.5	3	1.5	4.5	7

En el Cuadro A1 identificamos el valor más pequeño, $X=2$, a este valor le asignaremos el rango 1 [$R(2)=1$], Cuadro B1; el segundo valor más pequeño es $X=3$, Cuadro A1, pero en esta ocasión encontramos dos valores de X que son iguales a 3, en condiciones normales les deberíamos asignar los rangos 2 y 3 indistintamente, pero en lugar de eso lo que hacemos es sumar los rangos y asignar el promedio a los dos valores [$(2+3)/2=2.5$], por tanto $R(3)=2.5$ en ambos casos, Cuadro B1; el siguiente valor de X al que le asignaremos rango es $X=5$, Cuadro A1, nuevamente

hay dos valores de X iguales a 5, por tanto procedemos de manera similar al caso anterior, sumamos los rangos correspondientes y asignamos el promedio $[(4+5)/2=4.5]$ entonces $R(5)=4.5$ en los dos casos, Cuadro B1; el siguiente valor $X=7$, Cuadro 1, recibe el rango 6 [$R(7)=6$] y finalmente al valor $X=10$, Cuadro A1, le corresponde el rango 7, [$R(10)=7$], Cuadro B1; para obtener todos los rangos de Y se procedió de manera análoga a lo expuesto con X , el resultado final se aprecia en el Cuadro B1. Para otorgar rangos cuando hay empates también podemos hacerlo de manera inversa a como lo hicimos en este ejemplo, es decir, comenzamos desde los valores más grandes de X y terminamos con los más pequeños, cuando los rangos coinciden asignamos el promedio; hacemos lo mismo con Y . Nuevamente el procedimiento elegido no importa, siempre y cuando el criterio de asignación de los rangos sea el mismo tanto para X como para Y .

El valor ρ pertenece a la estadística no paramétrica⁶⁰, es decir, podemos aplicar ciertos métodos para probar hipótesis sobre él sin necesidad de tener supuestos acerca del comportamiento de las variables que lo componen. En la siguiente sección hablaremos acerca de la prueba de hipótesis para el valor ρ .

Prueba de Hipótesis

La ρ de Spearman es utilizada para probar la independencia entre dos variables aleatorias (poblaciones).

Se tienen los siguientes casos:

- **Caso A:** Prueba de dos colas

H₀: X y Y son independientes.

⁶⁰ La estadística no paramétrica es aquella en la que se emplean métodos para los cuales no hay necesidad de tener suposiciones acerca de las variables de estudio, suposiciones tales como: escalas intervalares, distribución normal, grandes muestras, homogeneidad de las varianzas, etc. Dichos métodos reciben el nombre de métodos no paramétricos debido a que no prueban valores paramétricos de tipos de población conocidos.

Ha: Existe una tendencia para que los valores más grandes de **X** estén apareados con los valores más grandes de **Y** y los valores más chicos de **X** estén apareados con los valores más chicos de **Y** ($\rho > 0$) ó existe una tendencia para que los valores más grandes de **X** estén apareados con los valores más chicos de **Y** y los valores chicos de **X** estén apareados con los valores grandes de **Y** ($\rho < 0$).

• **Caso B:** Prueba de una cola.

Ho: **X** y **Y** son independientes.

Ha: Existe una tendencia para que los valores más grandes de **X** estén apareados con los valores más grandes de **Y** y los valores más chicos de **X** estén apareados con los valores más chicos de **Y** ($\rho > 0$).

• **Caso C:** Prueba de una cola.

Ho: **X** y **Y** son independientes.

Ha: Existe una tendencia para que los valores más grandes de **X** estén apareados con los valores más chicos de **Y** y los valores más chicos de **X** estén apareados con los valores más grandes de **Y** ($\rho < 0$).

Usando el valor de ρ como estadístico⁶¹ de prueba tenemos la siguiente regla de decisión para cada uno de los casos antes citados:

Se rechaza **Ho** a nivel de significancia α si:

Caso A: $\rho < w_{\alpha/2}$ ó $\rho > w_{1-\alpha/2}$.

Caso B: $\rho > w_{1-\alpha}$.

Caso C: $\rho < w_{\alpha}$.

⁶¹ El estadístico es un valor que sirve como estimación de un parámetro para la población de estudio.

Donde w_i es el cuantil⁶² w a nivel de significancia i de la distribución p de Pearson.

Para ver los valores de los cuantiles W_i debemos consultar la **TABLA X** que se encuentra al final de este apéndice.

Ejemplo

A fin de poner en práctica todo lo expuesto acerca del estadístico p procederemos con el siguiente ejemplo:⁶³

Tomemos en cuenta las calificaciones que obtuvieron los alumnos de cierta escuela en la prueba de aptitud para las matemáticas (variable **X**) y en el examen departamental de aritmética (variable **Y**). La información se encuentra vertida en las dos primeras columnas de la **TABLA 2**. De la misma manera en que lo hicimos con el coeficiente Y , procedemos a graficar los datos de las columnas I y II para establecer si podría o no existir alguna relación lineal entre las variables en cuestión.

⁶² El cuantil se maneja como el valor crítico que debe rebasar el estadístico de prueba a fin de rechazar la hipótesis nula (H_0).

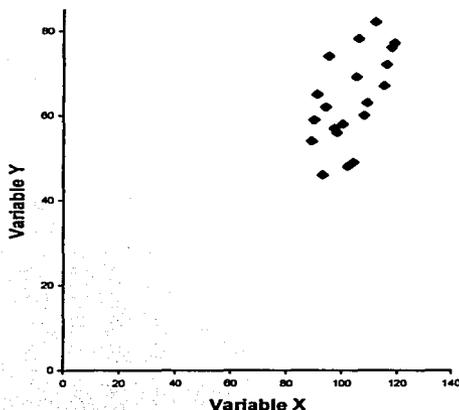
⁶³ Gronlud, N. p. I

TABLA 2

Nombre del Alumno	I Resultados prueba de aptitud. X	II Resultados prueba de aritmética Y	III Lugar en aptitud. Rango X. R(x)	IV Lugar en aritmética. Rango Y. R(y)	V Diferencia en lugar. (R _x)-R _(y) (d _i)	VI Cuadrado de la diferencia. (d _i) ²
Juan	119	77	1	3	-2	4
Enrique	118	76	2	4	-2	4
María	116	72	3	6	-3	9
Susana	115	67	4	8	-4	16
Guillermo	112	82	5	1	4	16
Carlos	109	63	6	10	-4	16
Engracia	108	60	7	12	-5	25
Rafael	106	78	8	2	6	36
Juana	105	69	9	7	2	4
Carlota	104	49	10	18	-8	64
Jaime	102	48	11	19	-8	64
Francisco	100	58	12	14	-2	4
Carmen	98	56	13	16	-3	9
Juanita	97	57	14	15	-1	1
Rubi	95	74	15	5	10	100
Edna	94	62	16	11	5	25
Elena	93	46	17	20	-3	9
Jorge	91	65	18	9	9	81
Alicia	90	59	19	13	6	36
Martín	89	54	20	17	3	9
						$\Sigma d_i = 534$

Un pequeño análisis de la dispersión de los puntos en la Gráfica 2 nos muestra cierta propensión de estos a ajustarse a una línea recta, los valores pequeños de X tienden a asociarse con los valores pequeños de Y y de igual forma los valores grandes de X tienden a aparearse con los valores grandes de Y.

Gráfica 2
Gráfica de Dispersión



Una vez revisada toda la información, podemos proseguir a calcular las siguientes columnas de la TABLA 2, lo primero que hay que hacer es identificar los rangos de los datos tanto de X como de Y (columnas III y IV respectivamente) para posteriormente calcular la diferencia d_i entre ellos (columna V), elevar al cuadrado este último resultado (columna VI) y por último sumar todas las d_i (desde $i=1$ hasta $n=20$). Los valores calculados son sustituidos en la expresión (5), obteniendo lo siguiente:

$$\rho = 1 - 6 \left(\frac{534}{20(20^2 - 1)} \right)$$

o bien:

$$\rho = .5984.$$

Este valor, en principio, podría parecer pequeño, lo que haremos a continuación es llevar a cabo una prueba de hipótesis.

Formulación de la Hipótesis Nula (H_0). Por tratarse de la ρ de Spearman, la hipótesis nula establecerá que las variables **X** y **Y** son independientes, es decir, no hay relación alguna entre la calificación de un alumno en la prueba de aptitud para las matemáticas (**X**) y su calificación en el examen departamental de aritmética (**Y**).

Selección del nivel de significancia. Para este ejemplo emplearemos el nivel de significancia de mayor uso: $\alpha=.05$, de tal forma que si al final de nuestra prueba rechazamos la hipótesis nula, la probabilidad de habernos equivocado será tan sólo del 5%.

Determinación del estadístico de prueba. El estadístico de prueba empleado aquí es el valor calculado de ρ en este ejemplo, es decir, $\rho=.598$.

Determinación de los valores críticos del estadístico de prueba. Aquí establecemos el valor crítico que el estadístico de prueba deberá rebasar a fin de rechazar la hipótesis nula. Para obtenerlo revisamos la **TABLA X⁶⁴**, que es la que le corresponde a ρ , de acuerdo con nuestro nivel de significancia $\alpha=.05$, el valor crítico del estadístico, para una prueba de una cola (**Caso B**), con una $n=20$ es: $W_{1-.05;2} = W_{1-.25} = W_{.75} = .4451$.

Comparación del valor crítico y del valor observado. En este caso, el valor calculado $\rho=.5984$ es mayor que $.4451 = W_{.75}$, debido a esto caemos en la región de rechazo y por tanto rechazamos H_0 a nivel de significancia $\alpha=.05$, o lo que es lo mismo, las variables **X** y **Y** no son independientes y más aún, se concluye que existe una relación *lineal positiva* entre las puntuaciones obtenidas en el examen de aptitud para las matemáticas y las conseguidas en el examen departamental de matemáticas.

⁶⁴ La tabla X se encuentra al final de este apéndice.

El ejemplo anterior es sumamente ilustrativo, ya que además de permitirnos explicar el cálculo del coeficiente ρ y su prueba de hipótesis, es en sí mismo un ejemplo de cómo validar un instrumento empleando el método de **validez predictiva**⁶⁵, con toda la información recabada estamos en condiciones de aseverar que si un alumno obtiene una valoración alta en el examen de aptitud entonces logrará acreditar de manera satisfactoria el examen departamental de aritmética.⁶⁶

¿Qué medida de correlación emplear?

A continuación haremos algunas observaciones:

- a) La Υ de Pearson, al igual que la ρ de Spearman, posee una prueba de hipótesis, sin embargo, para poder aplicarla, debemos pedir que los datos de las variables con los cuales se calcula su valor provengan de una población normal bivariada y cuando menos debemos contar con 30 observaciones de la forma (x,y); como estos supuestos no siempre se cumplen resulta más cómodo emplear la medida ρ .
- b) El procedimiento para la obtención de la fórmula que nos permite calcular Υ esta íntimamente ligado con el "Análisis de Regresión", esto quiere decir que Υ posee algunas características adicionales a las expuestas en estas notas.
- c) "Cuando se trabaja con datos a los que es aplicable tanto la Υ de Pearson como la ρ de Spearman, esta última tiene eficiencia del 91%; esto es, ρ es aproximadamente una prueba de la existencia de asociación entre dos variables en una población normal bivariada con una muestra de 100 casos tan sensible como la Υ de Pearson con 91 casos".⁶⁷

⁶⁵ Ver Capítulo III página 40.

⁶⁶ Esta última afirmación será cierta siempre y cuando las medidas obtenidas por la aplicación de los exámenes resulten ser **válidas y confiables**.

⁶⁷ Siegel Sydney, "Estadística no paramétrica" Ed. Trillas, México, 1994.

A final de cuentas las condiciones de nuestro problema determinarán cual será el coeficiente más adecuado, ya que en el fondo Υ y ρ poseen las mismas características.

TABLA X

Cuantiles de la prueba estadística de Spearman.

Prueba de una cola.						
N	P=.900	.950	.975	.990	.995	.999
4	.8000	.8000				
5	.7000	.8000	.9000	.9000		
6	.6000	.7714	.8286	.8857	.9429	
7	.5357	.6786	.7450	.8571	.8929	.9643
8	.5000	.6190	.7143	.8095	.8571	.9286
9	.4667	.5833	.6833	.7667	.8167	.9000
10	.4424	.5515	.6364	.7333	.7818	.8667
11	.4182	.5273	.6091	.7000	.7455	.8364
12	.3986	.4965	.5804	.6713	.7273	.8182
13	.3791	.4780	.5549	.6429	.6978	.7912
14	.3626	.4593	.5341	.6220	.6747	.7670
15	.3500	.4429	.5179	.6000	.6536	.7464
16	.3382	.4265	.5000	.5824	.6324	.7265
17	.3260	.4118	.4853	.5637	.6152	.7083
18	.3148	.3994	.4716	.5480	.5975	.6904
19	.3070	.3895	.4579	.5333	.5825	.6737
20	.2977	.3789	.4451	.5203	.5684	.6586
21	.2909	.3688	.4351	.5078	.5545	.6455
22	.2829	.3597	.4241	.4963	.5426	.6318
23	.2767	.3518	.4150	.4852	.5306	.6186
24	.2704	.3435	.4061	.4748	.5200	.6070
25	.2646	.3362	.3977	.4654	.5100	.5962
26	.2588	.3299	.3894	.4564	.5002	.5856
27	.2540	.3236	.3822	.4481	.4915	.5757
28	.2490	.3175	.3749	.4401	.4828	.5660
29	.2443	.3113	.3685	.4320	.4744	.5567
30	.2400	.3059	.3620	.4251	.4665	.5479

Para $n > 30$ la aproximación de los cuantiles para ρ se pueden obtener mediante la siguiente transformación:

$$W_p = \frac{x_p}{\sqrt{n-1}}$$

donde x_p es el cuantil p de una variable con distribución normal estandarizada este valor lo encontramos en la **Tabla XI**.

Los cuantiles pequeños, para pruebas de hipótesis de una cola en donde la H_a (Hipótesis alternativa) es $\rho < 0$, se obtienen a través de la ecuación:

$$W_p = -W_{1-p}$$

La región crítica (región de rechazo) corresponde a los valores de ρ más pequeños que (o más grandes que) el cuantil W_p (W_{1-p}), pero no incluyen el propio cuantil.

Nótese que la media de ρ es 0 (cero).

H_0 : (Hipótesis Nula): Las variables X y Y son independientes



APÉNDICE II

TEORÍA DE LOS TESTS.

La teoría de los tests se encarga, básicamente, de formular un modelo matemático teórico para medir la confiabilidad de un instrumento de medición.

La confiabilidad es la principal propiedad de interés a analizar, puesto que, al ser considerada como el *error aleatorio* (a diferencia del *error sistemático* o *validez*), ningún instrumento de medición construido queda exento de él.

En esta sección abordaremos las suposiciones que han hecho para generar toda esta teoría.

La medida.

Cada medida (puntaje) que obtenemos acerca de una variable que estamos observando consta de dos componentes:

- a) Una medida (puntaje) verdadero.
- b) Un error aleatorio.

Las aseveraciones anteriores se expresan en la siguiente ecuación:

$$t = T + e \quad (1)$$

Donde: t: puntaje observado.

 T: puntaje verdadero

 e: error aleatorio.

Las diferencias de los puntajes obtenidos en distintas aplicaciones de un mismo test (o en las diversas aplicaciones de tests *paralelos* o *equivalentes*⁶⁸) a un mismo individuo se deben al error aleatorio. El error aleatorio puede atribuirse a factores propios del individuo como lo son el cansancio, la ansiedad, el nerviosismo, la presión etc., o a factores ambientales que puedan distraer su atención; la presencia de este tipo de factores propiciarán que el puntaje observado cambie entre una y otra ejecución del test. A diferencia del puntaje observado el puntaje verdadero nunca cambia.

Definamos las variables que emplearemos de ahora en adelante.

Sean: e_i : error aleatorio del i -ésimo test.

T_i : puntaje verdadero obtenido en el i -ésimo test.

e : error aleatorio.

$E(e)$: esperanza del error aleatorio.

n : número de aplicaciones del test.

$Var(x)$: varianza de x

$Cov(x,y)$: covarianza de X y Y

$\gamma_{(j,k)}$: coeficiente de correlación de Pearson entre la variable j y la variable k .

Las suposiciones que se hacen entorno al error aleatorio (o error al azar) son las siguientes :

a) La esperanza de e es cero.

$$E(e) = 0$$

(2)

⁶⁸ La definición de instrumentos paralelos se discutirá mas adelante.

Lo anterior nos dice, entre otras cosas, que si sometiéramos a un individuo a la aplicación de un test un número infinito de veces o si aplicáramos un mismo test a un número infinito de individuos resultaría que en algunas ocasiones el error aleatorio (e_i) sería "positivo y en algunas otras "negativo", es decir, el puntaje observado se encontraría por encima o por debajo del puntaje real pero en el límite se esperaría que los errores se cancelaran unos a otros, obteniendo así una media (\bar{e}) o promedio igual a cero.

- b) La correlación entre el puntaje real y el error aleatorio es cero.

$$\gamma_{(T,e)} = 0 \quad (3)$$

- c) La correlación entre el error aleatorio de un test i y el puntaje real de otro test j es 0.

$$\gamma_{(e_i,T_j)} = 0 \quad \forall i \neq j \quad (4)$$

- d) El coeficiente de correlación entre los errores aleatorios de dos test distintos es cero.

$$\gamma_{(e_i,e_j)} = 0 \quad \forall i \neq j \quad (5)$$

De las propiedades del error aleatorio se desprenden dos resultados importantes (siempre y cuando apliquemos un test un número infinito de veces).

- a) La esperanza del puntaje observado y el puntaje real son iguales.

$$E(t) = E(T+e) = E(T) + E(e) = E(T) + 0 = E(T) \quad (6)$$

- b) La varianza del puntaje verdadero es equivalente a la suma de la varianza del puntaje real más la varianza del error aleatorio.

$$\begin{aligned}
\text{Var}(t) &= \text{Var}(T+e) = \text{Var}(T) + \text{Var}(e) + 2\text{Cov}(T,e) \\
&= \text{Var}(T) + \text{Var}(e) + 0 \\
&= \text{Var}(T) + \text{Var}(e)
\end{aligned}
\tag{7}$$

Tests paralelos.

Dos medidas son definidas como paralelas si ambas tienen idénticos puntajes verdaderos e iguales varianzas.

Si t y t' representan puntajes observados, con $t = T_1 + e_1$ y $t' = T_2 + e_2$ entonces t y t' serán paralelos si:

- a) Las varianzas de e_1 y e_2 son iguales : $\text{Var}(e_1) = \text{Var}(e_2)$
- b) Los puntajes verdaderos son los mismos: $T_1 = T_2$

Además de estas propiedades se adicionan otras⁶⁹ las cuales se listan a continuación:

- c) Las esperanzas (medias) de dos puntajes observados que sean paralelos serán las mismas: $E(t) = E(t')$
- d) Las varianzas de dos puntajes observados que sean paralelos serán las mismas: $\text{Var}(t) = \text{Var}(t')$
- e) Las intercorrelaciones de par en par entre puntajes observados paralelos son las mismas:

$$\gamma_{(tt')} = \gamma_{(tt')} = \gamma_{(tt')} \tag{8}$$

- f) La correlación de puntajes paralelos observados con otras variables, digamos la variable y , son las mismas:

⁶⁹ Carmines, Edward G. "Reliability assessment" Beverly Hills : Sage, c1979.

$$\gamma_{(ty)} = \gamma_{(ty)} = \gamma_{(t,y)} \quad (9)$$

Estos supuestos nos dicen que las diferencias que existen entre dos test que son considerados paralelos se deben únicamente a fluctuaciones aleatorias (error aleatorio).

Ítems Paralelos

Si pensamos en dos ítems como si cada uno de ellos fuera un test y los test fuesen paralelos, podemos concluir que dos ítems serán paralelos si sus medias, varianzas e intercorrelaciones son las mismas.

La confiabilidad.

Como se ha mencionado antes la confiabilidad se refiere a la consistencia de las medidas que se obtienen en distintas aplicaciones de un test, si pudiéramos eliminar todas las huellas (factor memoria) de la primera aplicación de un test podríamos aplicarlo a diferentes tiempos obteniendo diferentes puntuaciones para posteriormente calcular su confiabilidad por medio de un coeficiente de correlación, sin embargo, existe un método alternativo para estimar la confiabilidad y consiste en construir dos test que sean paralelos.

Si construyéramos dos instrumentos que fuesen paralelos, pensemos en t_i y t_j con $t_i = T_i + e_i$ donde t_i representa la puntuación verdadera i -ésimo individuo y e_i el error aleatorio del i -ésimo individuo (análogamente para t_j) podríamos calcular el coeficiente de correlación de Pearson entre estas medidas o puntajes de la siguiente manera:

$$\gamma_{(t_i, t_j)} = \frac{\text{Cov}(t_i, t_j)}{\sqrt{\text{Var}(t_i)\text{Var}(t_j)}} = \frac{\text{Cov}(T_i + e_i, T_j + e_j)}{\sqrt{\text{Var}(t_i)\text{Var}(t_j)}}$$

$$\gamma_{(t_i, t_j)} = \frac{\text{Cov}(T_i, T_j) + \text{Cov}(T_j, e_i) + \text{Cov}(T_i, e_j) + \text{Cov}(e_i, e_j)}{\sqrt{\text{Var}(t_i)^2}} = \frac{\text{Cov}(T_i, T_j)}{\sqrt{\text{Var}(t_i)^2}}$$

Dado que hemos supuesto que t_i y t_j son equivalentes podemos eliminar los subíndices obteniendo finalmente.

$$\gamma_{(t,t)} = \frac{\text{Var}(T)}{\text{Var}(t)} \quad (10)$$

De la ecuación (7) sabemos que : $\text{Var}(T) = \text{Var}(t) - \text{VAR}(e)$

Por lo tanto:

$$\gamma_{(t,t)} = \frac{\text{Var}(t) - \text{Var}(e)}{\text{Var}(t)} = 1 - \frac{\text{Var}(e)}{\text{Var}(t)} \quad (11)$$

La ecuación (11) es la definición clásica de la confiabilidad.

La Confiabilidad y la longitud del test.

La confiabilidad de un test está en función del número de ítems que lo componen, de acuerdo con la ecuación (11) tenemos que :

$$\gamma_{(t,t)} = \frac{\text{Var}(T)}{\text{Var}(t)}$$

En esta ecuación se aprecia fácilmente que la confiabilidad es la proporción de la varianza verdadera de la varianza total. A continuación examinaremos cómo la varianza total ($\text{Var}[t]$), y la varianza verdadera ($\text{Var}[T]$), son afectadas cuando se aumenta la longitud de un test. Definimos la longitud de un test como el número de ítems que lo conforman

Varianza de un test.

La varianza de un test es la suma de las varianzas de cada uno de los ítems que lo componen, de tal forma que si tenemos un test t_i con n ítems la varianza queda determinada por:

$$\begin{aligned}
S^2(t_i) &= \sum_{j=1}^n S^2(x_{t_{ij}}) = S^2_{i1} + S^2_{i2} + S^2_{i3} + \dots + S^2_{in} + 2\text{Cov}(x_{i1}, x_{i2}) + \\
&+ 2\text{Cov}(x_{i1}, x_{i3}) + \dots + 2\text{Cov}(x_{i1}, x_{ij}) + \dots + 2\text{Cov}(x_{i1}, x_{in}) + \\
&+ 2\text{Cov}(x_{i(n-1)}, x_{in})
\end{aligned} \tag{12}$$

Donde : $S^2(t_i)$ = Varianza total del test t_i

$x_{t_{ij}}$ = Puntuación del j -ésimo ítem del test t_i

$S^2(x_{t_{ij}})$ = Varianza de $x_{t_{ij}}$

$\text{Cov}(x_{t_{ij}}, x_{t_{ik}})$ = Covarianza de $x_{t_{ij}}$ y $x_{t_{ik}}$ con $j \neq k$

Si recordamos que el coeficiente de correlación de Pearson entre dos variables, digamos $x_{t_{ij}}$ y $x_{t_{ik}}$ con $j \neq k$ es :

$$\gamma(x_{t_{ij}}, x_{t_{ik}}) = \frac{\text{Cov}(x_{t_{ij}}, x_{t_{ik}})}{S_{t_{ij}} S_{t_{ik}}} \tag{13}$$

Donde : $\gamma(x_{t_{ij}}, x_{t_{ik}})$ = Coeficiente de correlación de $x_{t_{ij}}$ y $x_{t_{ik}}$

$\text{Cov}(x_{t_{ij}}, x_{t_{ik}})$ = Covarianza de $x_{t_{ij}}$ y $x_{t_{ik}}$

$S_{t_{ij}}$ = Desviación estándar del j -ésimo ítem del test t_i

y agrupamos los términos de la expresión (12) obtendremos:

$$S^2(t_i) = \sum_{j=1}^n S^2_{t_{ij}} + 2 \sum_{j=1}^{n-1} \gamma(x_{t_{ij}}, x_{t_{ik}}) S_{t_{ij}} S_{t_{ik}} \tag{14}$$

la cual es la varianza del test t_i

Varianza total.

Consideremos el caso de n test paralelos, cada uno de ellos con una varianza $S^2(t_i)$ para $i = 1, 2, 3, \dots, n$ tal y como se definió en la sección anterior, pongamos los puntajes crudos en renglones y columnas, donde g y h representan dos test cualesquiera de la matriz. Para cada renglón, es decir, para cada test, obtenemos entonces los términos de varianza mostrados en la Tabla I

TABLA I

Test	Suma de los términos de varianza
1	$S^2(t_1) + \sum_{k=1}^n \gamma(t_1, t_k) S_{1k} S_{1k} \quad \forall k \neq 1$
2	$S^2(t_2) + \sum_{k=1}^n \gamma(t_2, t_k) S_{2k} S_{2k} \quad \forall k \neq 2$
3	$S^2(t_3) + \sum_{k=1}^n \gamma(t_3, t_k) S_{3k} S_{3k} \quad \forall k \neq 3$
...	
g	$S^2(t_g) + \sum_{k=1}^n \gamma(t_g, t_k) S_{gk} S_{gk} \quad \forall k \neq g$
...	
n	$S^2(t_n) + \sum_{k=1}^n \gamma(t_n, t_k) S_{nk} S_{nk} \quad \forall k \neq n$

La suma total de los términos de varianza de un test compuesto de n tests paralelos S^2_{nt} es:

$$S^2_{nt} = \sum_{i=1}^n S^2(t_i) + \sum_{i=1}^n \sum_{j=1}^n \gamma(t_i, t_j) S_{ij} S_{ij} \quad \forall i \neq j \quad (15)$$

Donde : S^2_{nt} = Varianza total un test compuesto de n test paralelos.

$S^2(t_i)$ = Varianza total del test t_i

$\gamma(t_i, t_j)$ = Coeficiente de correlación de t_i y t_j

S_{t_i} = Desviación estándar de t_i

Debido a que las varianzas de los test paralelos son iguales la primera suma de la expresión (15) queda como:

$$\sum_{i=1}^n S^2(t_i) = nS^2(t) \quad \text{con } t = t_g \quad \text{para algún } g \in \{1, 2, 3, \dots, n\} \quad (16)$$

La correlación $\gamma(t_i, t_j)$, que se encuentra en todos los términos de la segunda suma de la expresión (15) es el coeficiente de correlación para la relación entre test paralelos y es el mismo para cada sumando, denotémoslo por $\gamma(t, t)$. Las desviaciones estándar son iguales para test paralelos, $S_{t_i} = S_{t_j}$ para $i \neq j$, por lo tanto $S_{t_i} S_{t_j} = S^2(t)$. Dado que tenemos $n(n-1)$ términos en la segunda suma y todos son iguales obtenemos:

$$\sum_{i=1}^n \sum_{j=1}^n \gamma(t_i, t_j) S_{t_i} S_{t_j} = n(n-1) \gamma(t, t) S^2(t) \quad (17)$$

Sustituyendo (16) y (17) en (15) la ecuación para el cálculo de la varianza total de n test paralelos puede escribirse ahora como:

$$\begin{aligned} S^2_{\pi} &= nS^2(t) + n(n-1)\gamma(t, t)S^2(t) \\ S^2_{\pi} &= nS^2(t) [1 + (n-1)\gamma(t, t)] \end{aligned} \quad (18)$$

Esta es la varianza total del test t compuesto de n test paralelos, es decir, cuando la longitud del test se ha aumentado n veces.

Varianza verdadera.

El análisis efectuado para establecer cual es la varianza de los puntajes verdaderos (varianza verdadera) es análogo al que realizamos en la sección anterior

para el cálculo de la varianza total, si denotamos S^2_{nT} como la varianza de los puntajes verdaderos para n test paralelos tendremos que :

$$S^2_{nT} = \sum_{i=1}^n S^2(T_i) + \sum_{i=1}^n \sum_{j=1}^n \gamma(T_i, T_j) S_{T_i} S_{T_j} \quad \forall i \neq j \quad (19)$$

Puesto que la varianza verdadera es la misma para todos los test paralelos el primer término de la ecuación (19) es :

$$\sum_{i=1}^n S^2(T_i) = n S^2(T) \quad \text{con } T = T_g \quad \text{para algún } g \in \{1, 2, 3, \dots, n\} \quad (20)$$

La correlación entre los puntajes verdaderos de tests paralelos es 1, por consiguiente todos los términos de la forma $\gamma(T_i, T_j)$, serán constantes e iguales a 1. Como las desviaciones estándar de los puntajes verdaderos son las mismas para tests paralelos, los productos $S_{T_i} S_{T_j}$ para $i \neq j$ serán iguales y los denotaremos por S^2_T . Por último, en el segundo término de la ecuación (19) tenemos $n(n-1)$ sumandos y todos son iguales. La expresión final después de simplificar es :

$$\sum_{i=1}^n \sum_{j=1}^n \gamma(T_i, T_j) S_{T_i} S_{T_j} = n(n-1) S^2(T) \quad (21)$$

Sustituyendo (20) y (21) en (19) la ecuación para el cálculo de la varianza verdadera de n test paralelos puede escribirse ahora como:

$$S^2_{nT} = n S^2(T) + n(n-1) S^2(T)$$

$$S^2_{nT} = n^2 S^2(T) \quad (22)$$

La ecuación (22) nos dice que la varianza verdadera se incrementa como el cuadrado de n cuando la longitud de un test se aumenta n veces.

Una vez que hemos analizado el comportamiento de la varianza total y la varianza verdadera cuando aumentamos la longitud de un test n veces, nos preguntamos que sucede con la confiabilidad de un test cuando aumentamos su longitud también n veces.

De la ecuación :

$$\gamma_{(t,t)} = \frac{S^2(T)}{S^2(t)}$$

Podemos derivar la confiabilidad para un test que ha aumentado su longitud n veces:

$$\gamma_{(t,tn)} = \frac{S^2_n(T)}{S^2_n(t)}$$

que de acuerdo a las ecuaciones (10), (18) y (22) obtenemos:

$$\gamma_{(t,tn)} = \frac{S^2_n(T)}{S^2_n(t)} = \frac{n^2 S^2(T)}{n S^2(t) [1 + (n-1) \gamma(t,t)]}$$

$$\gamma_{(t,tn)} = \frac{n \gamma(t,t)}{1 + (n-1) \gamma(t,t)} \quad (23)$$

donde n es el número de veces que el test se aumentó en longitud, $\gamma(t,tn)$ es la confiabilidad del test al aumentar su longitud n veces y $\gamma(t,t)$ es confiabilidad del test inicial.

Esta expresión es conocida comúnmente como la fórmula de profética de **Spearman-Brown** y para el caso en que $n = 2$ tenemos:

$$\gamma_{(t,2t)} = \frac{2\gamma(t,t)}{1 + \gamma(t,t)} \quad (24)$$

que es la fórmula que empleamos para calcular la confiabilidad con el método llamado "División por mitades".

Método de Kuder-Richardson y α de Cronbach.

El método de División por mitades o Formas Paralelas esta basado en el hecho de que podemos dividir un test en dos partes, donde cada una de ellas es equivalente de la otra de tal forma que podemos calcular la confiabilidad total a partir del procedimiento ya expuesto.

Kuder y Richardson derivaron algunas ecuaciones para calcular la confiabilidad de un test sin necesidad de partirlo por la mitad, suponiendo que cada ítem representa un test paralelo, en base a esta suposición podemos decir todos los ítems tienen la **misma media** y la **misma varianza**.

De la ecuación (11) recordamos que la expresión que nos permite calcular la varianza de un test compuesto de n ítems es :

$$S^2(t) = \sum_{j=1}^n S_i^2 + 2 \sum_{j=1}^{n-1} \gamma(x_{ik}, x_{ij}) S_{ij} S_{ik}$$

De resultados anteriores sabemos de antemano que hay $n(n-1)$ términos incluidos en el segundo término de la expresión de arriba, si suponemos que cada ítem es paralelo a cada uno de los otros ítems tendremos que cada uno de ellos tiene la misma frecuencia de respuestas correctas y la misma correlación con los otros ítems, podemos entonces reemplazar $S_i S_k$ por S_i^2 y la suma de los términos de covarianza puede escribirse como $n(n-1)\gamma(x_i, x_k) S_i^2$ obteniendo así:

$$S^2(t) = \sum_{j=1}^n S_i^2 + n(n-1)\gamma(x_i, x_k) S_i^2 \quad (25)$$

si resolvemos esta última ecuación para $\gamma(x_i, x_k)$ con $n\bar{s}_i^2 = \sum S_i^2$ obtenemos :

$$\gamma(x_{i_1}, x_{i_k}) = \frac{S^2(t) - \sum_{j=1}^n S^2 i}{(n-1) \sum_{j=1}^n S^2 i} \quad (26)$$

Esta es una expresión para la correlación promedio entre los ítems individuales del test.

Dado que $\gamma(x_i, x_k)$ da la correlación entre un ítem y su ítem paralelo, la ecuación 26 da también la confiabilidad de un solo ítem. Con la fórmula general de Spearman-Brown, podemos estimar la confiabilidad de un test compuesto de n ítems; recordando la expresión (23) tenemos :

$$\gamma_{(t,tn)} = \frac{n \gamma(t, t)}{1 + (n-1) \gamma(t, t)}$$

y reemplazando el valor de $\gamma(t, t_k)$ por el valor de $\gamma(x_i, x_k)$ de la ecuación (26) dado que asumimos que cada ítem representa un test paralelo obtenemos :

$$\gamma_{(t,tn)} = \frac{n \left[\frac{S^2(t) - \sum_{j=1}^n S^2 i}{(n-1) \sum_{j=1}^n S^2 i} \right]}{1 + (n-1) \left[\frac{S^2(t) - \sum_{j=1}^n S^2 i}{(n-1) \sum_{j=1}^n S^2 i} \right]} \quad (27)$$

Realizando las operaciones correspondientes podemos simplificar la expresión anterior obteniendo:

$$\gamma_{(t, t_n)} = \frac{n}{1 + (n-1)} \left[\frac{S^2(t) - \sum_{j=1}^n S^2 i}{S^2(t)} \right] \quad (28)$$

donde $\gamma(t, t_n)$ es conocido como el α de Cronbach y en el caso particular en que los ítems se clasifiquen como 0 o 1 la expresión adquiere la siguiente forma :

$$\gamma_{(t, t_n)} = \frac{n}{1 + (n-1)} \left[\frac{S^2(t) - \sum_{j=1}^n p_i q_i}{S^2(t)} \right] \quad (29)$$

Donde :

p_i : proporción de las personas que obtuvieron un 1 en el i -ésimo ítem del instrumento.

q_i : proporción de las personas que obtuvieron un 0 en el i -ésimo ítem del instrumento ($q_i = 1 - p_i$).

La ecuación (29) es conocida como el coeficiente de Kuder Richardson KR20.

Las propiedades de cada uno de los coeficientes que se han mencionado en esta sección es tema de estudio en el Capítulo III de esta tesis, aquí simplemente hemos desarrollado los supuestos que han dado origen a las fórmulas que requerimos para calcular la confiabilidad de un instrumento.

BIBLIOGRAFÍA

- Ander-Egg, Ezequiel "Técnicas de investigación social" Ed. El Ateneo. 21 Edición. México, 1989.
- Adkins Wood, Dorothy. "Elaboración de Tests. Desarrollo e interpretación de los tests de aprovechamiento". Ed. Trillas. 2a Edición. México, 1990 (Reimp. 2000).
- Atilano, López Enrique "Tesis: La Administración de la DGSCA", (Seminario de Investigación Administrativa), UNAM; Facultad de Contaduría y Administración, 1990.
- Carmines, Edward G. "Reliability assessment" Beverly Hills : Sage, c1979.
- Dunnette, Marvin "Psicología Industrial" Ed. Trillas. 2ª Edición. México, 1989 (Reimp. 1999).
- G. Carmines Edward, McIver P. Jhon. "Unidimensional Scaling" Indiana University Sage Publications.
- Gorec, Klaus- y otros. "Cálculo para evaluar la validez y confiabilidad para evaluar pruebas diagnósticas: Un estudio preliminar", Revista intercontinental de psicología y Educación, 1º de Junio año 1988.
- Gronlud, Norman Edward "Medición y evaluación en la enseñanza". Ed. Trillas. México, 1974.
- Gronlud, Norman Edward "Elaboración de Tests de Aprovechamiento". Ed. Trillas. México, 1974 (Reimp. 1999).
- Hernández, Roberto; Fernández, Carlos y Baptista, Pilar "Metodología de la investigación" Ed. Mc Graw Hill, México, 1996.
- Kinnear, Tomás C. "Investigación de Mercados: Un Enfoque Aplicado". Ed. Mc Graw Hill. México, 1993.

- Magnusson, David "Teoría de los tests: Psicología diferencial. Psicología aplicada. Orientación Vocacional". Ed. Trillas. 2ª Edición. México, 1990 (Reimp. 2000).
- Martínez-Guerrero, J. & Sánchez-Sosa, J. J. "Intervalidación social de estrategias docentes en la Facultad de Psicología". Métodos Docentes, UNAM 3, México, 1981.
- Morales, María Luisa. "Psicometría Aplicada". Ed. Trillas. 2ª Edición. México, 1990 (Reimp. 2002).
- Norville M., Downie "Métodos estadísticos aplicados", Ed. Harla S.A. de C.V. 5ª Edición. México, 1986.
- Paul, Hoel "Estadística Elemental" Ed. Cia Editorial Continental S.A. de C.V. 4ª Edición. México, 1975.
- Sánchez Hidalgo, Efraín. "Psicología Educativa" Ed. Universitaria. Universidad de Puerto Rico, 7a Edición, 1972.
- Siegel Sydney, "Estadística no paramétrica" Ed. Trillas, México, 1994.
- Rodríguez, Mauro y Austria, Honorata "Formación de Instructores" Ed. Mc Graw Hill, 2ª Edición, México, 1999.
- Thurstone, Louis L. "La medición de la inteligencia; la aptitud y el interés", Ed. Paidós, Buenos Aires, 1967.
- Weiers, Ronald M. "Investigación de Mercados". Ed Prentice Hall-Hispanoamericana S.A. México, 1986.

DIRECCIONES EN INTERNET

- Asociación Mexicana de Educación Continua (AMEC).

<http://www.unam.mx/redec/Congreso/AMEC/Amecindex.html>

- Dirección General de Servicios de Cómputo Académico (DGSCA).

<http://www.dgsca.unam.mx/dgsca.html>

- Tecnológico de Antioquia.

<http://www.tdea.edu.co/>

- Universidad de Guadalajara.

<http://galeon.hispavista.com/escuela/investigacion.html>

DIRECCIONES EN INTERNET

- Asociación Mexicana de Educación Continua (AMEC).

<http://www.unam.mx/redec/Congreso/AMEC/Amecindex.html>

- Dirección General de Servicios de Cómputo Académico (DGSCA).

<http://www.dgsca.unam.mx/dgsca.html>

- Tecnológico de Antioquia.

<http://www.tdea.edu.co/>

- Universidad de Guadalajara.

<http://galeon.hispavista.com/escuela/investigacion.html>