

IPN

INSTITUTO POLITECNICO NACIONAL

**ESCUELA SUPERIOR DE FISICA Y
MATEMATICAS**

**MODELO DE SELECCIÓN DE LOS MEJORES
CANDIDATOS PARA UN EMPLEO.**

AUTOR: GERMAN MORENO URBANO

SEMINARIO DE TITULACION

**“MÉTODOS ESTADÍSTICOS UNIVARIADOS Y
MULTIVARIADOS APLICADOS AL ANÁLISIS DE
DATOS”**

MEXICO D.F. A 14 DE AGOSTO DEL 2006

INDICE.

UNIDAD I.-INTRODUCCION

1.- Introducción.....	5
1.1.- Problema.....	5
1.2.-Justificación.....	5
1.3.- Objetivo.....	6
1.4.- Hipótesis.....	6

UNIDAD II.- CONTEXTO INSTITUCIONAL

2.- Contexto Institucional.....	7
2.1.- Instituto Politécnico Nacional.....	7
2.1.1.- Misión.....	8
2.1.2.- Visión.....	8
2.1.3.- Finalidades.....	10
2.1.4.- Estructura Orgánica.....	11
2.2.- Descripción de las características de la escuela.....	13
2.2.1.- Historia.....	13
2.3.- Características de la Carrera de Ingeniería.....	14
2.3.1.- Objetivo de la carrera.....	14
2.3.2.- Perfil del alumno.....	14
2.3.2.1.- Ingreso.....	10
2.3.3.- Perfil de egreso.....	15
2.3.3.1.-Conocimientos.....	15
2.3.3.2.- Habilidades.....	15

2.3.3.3.- Actitudes y valores.....	16
------------------------------------	----

UNIDAD III.- FUNDAMENTOS TEORICOS

3.-Fundamentos teóricos.....	17
3.1 Perfil de puesto.....	17
3.2.-Elección de las técnicas de selección.....	19
3.2.1.-Entrevista de selección.....	19
3.2.2.-Pruebas de conocimiento o de capacidad.....	22
3.2.3.-Pruebas psicométricas.....	23
3.2.4.-Pruebas de personalidad.....	26
3.2.5.-Técnicas de simulación.....	27
3.3.- Análisis Multivariado.....	28
3.3.1.-Objetivos de los Métodos Multivariados.....	29
3.3.2.-Clasificación de los Métodos Multivariados.....	30
3.3.3.-Notación de matrices y vectores.....	30
3.3.3.1.- Matriz de datos.....	31
3.3.3.2.-Vectores de datos.....	31
3.3.4.- Esperanzas y varianzas de vectores aleatorios.....	32
3.3.4.1.-Media.....	32
3.3.4.2.-Varianzas-Covarianzas.....	32
3.3.4.3.-Correlaciones.....	33
3.3.4.4.-Comentarios.....	34
3.4.-Análisis exploratorio multivariado.....	35
3.4.1-Estadísticas multivariadas descriptivas.....	35

3.4.1.1.- Media muestral.....	36
3.4.1.2.- Varianza muestral.....	36
3.4.1.3.- Correlación muestral.....	37
3.4.2.-Análisis gráfico de datos multivariados.....	38
3.4.2.1.-Diagramas de dispersión (bidimensional).....	38
3.4.2.2.- Diagramas de dispersión (tridimensional).....	38
3.4.3.- Análisis de componentes principales.....	39
3.4.3.1.-Breve repaso de Matrices.....	39
3.4.4.- Componentes principales.....	43
3.4.4.1.- Objetivos de las componentes principales.....	43
3.4.4.2.- Características.....	44
3.4.4.3.- Componentes.....	44
3.4.4.4- Interpretación de λ_k	45
3.4.4.5.- Valores o marcadores.....	47
3.4.5.-Componentes principales sobre variables estandarizadas.....	47
UNIDAD IV.- DESARROLLO DEL PROBLEMA	
4.1.- Desarrollo del problema.....	49
4.2.-Conclusiones y sugerencias.....	62
ANEXOS.....	63

UNIDAD I

1.- INTRODUCCION

En la actualidad, el hecho de elegir al personal adecuado para los distintos puestos que una empresa ofrece, requiere de una selección minuciosa de las características de los aspirantes, de acuerdo a los requerimientos que se tengan para dicho puesto. Estas características se le conocen como Perfil de Puesto. La importancia de la selección de personal varía según los objetivos de la empresa.

1.1-PROBLEMA

Se requiere encontrar a partir de variables predeterminadas a los mejores candidatos que puedan desempeñar de forma aceptable algún empleo.

Las variables para medir el desempeño de los empleados, deberán ayudarnos a elegir al mejor candidato para realizar las actividades que se requieren. Estas actividades tendrán que estar perfectamente identificadas para poder definir el perfil del mejor candidato.

1.2- JUSTIFICACIÓN

Los Directivos invierten más del 60% de su tiempo resolviendo problemas relacionados con su gente. ¿Por qué sucede esto? Por la falta de información acerca de por qué sus empleados se comportan de la forma cómo lo hacen y qué es lo que los motiva.

Como es bien sabido, los beneficios y la rentabilidad de una empresa se ven influidas significativamente por la calidad de gente que se tiene, lo productiva que sea y lo motivada que esté para lograr los objetivos que se les asignan. Por ello, las empresas necesitan ver el retorno de la inversión que efectúan en sus recursos humanos. De modo que los Directivos

deben centrar sus esfuerzos en lograr este retorno monetario a través de los resultados que consigan por medio de sus empleados. En la medida en que esto suceda en una Organización, la misma verá como se reduce considerablemente el porcentaje del tiempo que sus Directivos emplean en resolver problemas relacionados con su gente.

1.3- OBJETIVO

Determinar al mejor candidato para un puesto determinado, de acuerdo al perfil de dicho puesto.

1.4- HIPOTESIS

Existen variables que de alguna manera se pueden medir para poder determinar las características de personas y de las actividades que desarrollara en dentro del ámbito laboral. Dichas variable están relacionadas y nos pueden ayudar a determinar que persona es la mas capaz de cumplir de mejor forma dicha actividad.

Se espera que los resultados de este modelo, nos ayuden a realizar una mejor selección de personal y con eso reducir los costos que conllevan contratar a la persona equivocada para desarrollar las actividades asignadas.

UNIDAD II

2.- CONTEXTO INSTITUCIONAL

2.1- INSTITUTO POLITÉCNICO NACIONAL

El Instituto Politécnico Nacional estuvo asistido en su nacimiento por dos hombres visionarios, el Gral. Lázaro Cárdenas y el Ing. Juan Carlos de Diós Batíz, cuya conducta moral y actitud revolucionaria dictaron el código institucional -no escrito- de responsabilidad social e identidad.

En los años de 1935 y 1936 en que se planeaba la creación del I.P.N., fue fundamental la labor que realizó el Ing. Juan de Diós Batíz, pues gracias a su dinamismo y al interés del entonces presidente Lázaro Cárdenas, pudieron sentarse las bases que hicieron posible la creación del Politécnico y de su proyección y progreso continuos.

A 70 años de su creación el I.P.N. tiene como objetivo la formación académica, científica y cultural de jóvenes en forma integral y con un profundo sentido nacionalista, de servicio a la sociedad en la cual se encuentran inmersos, atendiendo fielmente la interpretación de su lema:

"LA TÉCNICA AL SERVICIO DE LA PATRIA"

2.1.1- MISIÓN

El Instituto Politécnico Nacional es la institución educativa laica, gratuita de Estado, rectora de la educación tecnológica pública en México, líder en la generación, aplicación, difusión y transferencia del conocimiento científico y tecnológico, creada para contribuir al desarrollo económico, social y político de la nación. Para lograrlo, su comunidad forma integralmente profesionales en los niveles medio superior, superior y posgrado, realiza investigación y extiende a la sociedad sus resultados, con calidad, responsabilidad, ética, tolerancia y compromiso social.

2.1.2- VISIÓN

Una institución educativa innovadora, flexible, centrada en el aprendizaje; fortalecida en su carácter rector de la educación pública tecnológica en México; poseedora de personalidad jurídica y patrimonio propios, con capacidad de gobernarse a sí misma; enfocada a la generación y difusión del conocimiento de calidad; caracterizada por procesos de gestión transparentes y eficientes; con reconocimiento social amplio por sus resultados y sus contribuciones al desarrollo nacional; por todo ello, posicionada estratégicamente en los ámbitos nacional e internacional.

Participa en el Sistema Educativo Nacional, comparte recursos intra y extra institucionales, intercambia información y conduce proyectos educativos y de investigación conjuntos, ubicando su operación en rangos de excelencia definidos por indicadores internacionales, constituyéndose en referentes del Sistema Nacional de Educación Científica y Tecnológica.

Cuenta con un sistema de educación virtual consolidado, con programas educativos y de formación a lo largo de la vida.

Tiene integrados sus distintos niveles formativos y las diferentes modalidades educativas. Posee una importante fortaleza en materia de uso de las tecnologías de información y de comunicación, aplicados a sus procesos académicos, de investigación y de extensión y difusión.

Sus procesos formativos, la integración de su planta docente y la investigación realizada, cumplen con normas de calidad definidas por instancias nacionales e internacionales. Los mecanismos de evaluación y la rendición de cuentas garantizan que su comunidad y la sociedad confirmen que la calidad es una constante en todas las acciones y procesos. Los programas académicos son acreditados y sus egresados cuentan con la certificación correspondiente.

Con un modelo de investigación basado en redes de cooperación nacional e internacional, plenamente vinculado con los sectores productivo y social, fomenta la generación, uso, circulación y protección del conocimiento en sectores estratégicos que promueven la competitividad, la equidad y el mejoramiento de la sociedad.

La comunidad del IPN tiene una conformación multinacional; su desempeño socialmente comprometido la habilita para actuar de manera proactiva, con capacidad para diseñar soluciones originales a problemas y oportunidades emergentes.

Cuenta con un modelo integral de vinculación, basado en programas académicos y de investigación que impulsan la incubación y desarrollo de la micro, pequeña y mediana empresa; el liderazgo social y empresarial de sus alumnos, garantizando la calidad de los servicios prestados a los sectores productivos.

Como institución rectora en la tecnología, posee instalaciones modernas, funcionales y equipadas con tecnología de punta que se utiliza de manera eficiente, eficaz y pertinente.

2.1.3- FINALIDADES

Son finalidades del Instituto Politécnico Nacional:

I.- Contribuir a través del proceso educativo a la transformación de la sociedad en un sentido Democrático, y de progreso social, para lograr la justa distribución de los bienes materiales y culturales dentro de un régimen de igualdad y libertad;

II.- Realizar investigación científica y tecnológica con vista al avance del conocimiento, al desarrollo de la enseñanza tecnológica y al mejor aprovechamiento social de los recursos naturales y materiales;

III.- Formar profesionales e investigadores en los diversos campos de la ciencia y la tecnología, de acuerdo con los requerimientos del desarrollo económico, político y social del país;

IV.- Coadyuvar a la preparación técnica de los trabajadores para su mejoramiento económico y social;

V.- Investigar, crear, conservar y difundir la cultura para fortalecer la conciencia de la nacionalidad,

procurar el desarrollo de un elevado sentido de convivencia humana y fomentar en los educandos el

amor a la paz y los sentimientos de solidaridad hacia los pueblos que luchan por su Independencia;

VI.- Promover en sus alumnos y egresados actitudes solidarias y democráticas que reafirmen nuestra independencia económica;

VII.- Garantizar y ampliar el acceso de estudiantes de escasos recursos a todos los servicios de la enseñanza técnica que preste el instituto;

VIII.- Participar en los programas que para coordinar las actividades de investigación se formulen de acuerdo con la planeación y desarrollo de la política nacional de ciencia y tecnología.

IX.- Contribuir a la planeación y al desarrollo interinstitucional de la Educación Técnica y realizar la función rectora de este tipo de Educación en el País, coordinándose con las demás instituciones que integran el Consejo del Sistema Nacional de Educación Tecnológica, en los términos previstos por la Ley para la Coordinación de la Educación Superior y de conformidad con los acuerdos que se tomen en el propio Consejo.

2.1.4- ESTRUCTURA ORGÁNICA

Son autoridades del Instituto:

I.- El Director General;

II.- El Secretario General;

III.- Los Secretarios de Area;

IV.- Los Directores de Coordinación, y

V.- Los Directores, Directores Adjuntos y Subdirectores de Escuelas, Centros y Unidades de Enseñanza y de Investigación.

Son órganos consultivos del Instituto:

- I.- El Consejo General Consultivo, y
- II.- Los Consejos Técnicos Consultivos Escolares.

Son órganos de apoyo dependientes del Instituto:

- I.- La Estación de Televisión XEIPN Canal Once del Distrito Federal;
- II.- El Centro Nacional de Cálculo.

Son organismos auxiliares del Instituto:

- I.- El Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional;
- II.- La Comisión de Operación y Fomento de Actividades Académicas del Instituto Politécnico Nacional;
- III.- El Patronato de Obras e Instalaciones del Instituto Politécnico Nacional.

2.2- DESCRIPCION DE LAS CARACTERÍSTICAS DE LA ESCUELA

2.2.1- HISTORIA

La Escuela Superior de Física y Matemáticas del Instituto Politécnico Nacional fue creada por decreto presidencial el 2 de marzo de 1961. Entre el año de su fundación y 1964 se redactaron sus planes de estudio y las condiciones académicas para obtener el título de Licenciado en Física y Matemáticas. En el mismo periodo se creó la Maestría en Ingeniería Nuclear.

Después, entre 1964 y 1971, fueron creadas las maestrías en Física, Matemáticas y Ciencias de los Materiales. En 1975 se fundó el Doctorado en Física.

Actualmente, la escuela imparte las carreras de: Licenciatura en Física y Matemáticas e Ingeniería Matemática, así como las maestrías en Física, Matemáticas, Ciencia de Materiales e Ingeniería Nuclear, y los doctorados en Física, Matemáticas y Física de Materiales.

Los egresados de la ESFM cuentan normalmente con una participación destacada en los campos de investigación y docencia. La preparación que reciben en la escuela les ha permitido desarrollarse no sólo en los diversos campos de la Física y las Matemáticas, sino también en áreas como la computación, la economía y las ciencias biológicas. Algunos de nuestros egresados también han incursionado exitosamente en la industria privada y en la administración pública.

2.3- CARACTERÍSTICAS DE LA CARRERA DE INGENIERÍA

2.3.1- OBJETIVO DE LA CARRERA

El objetivo de la carrera de Ingeniería Matemática es la formación de profesionistas que aplicarán o crearán los modelos industriales, administrativos y financieros. Modelos que han de ser la base para que profesionales de otras especialidades investiguen, diseñen, produzcan, construyan, operen, mantengan estimen y dirijan sistemas con diversos grados de complejidad. El egresado podrá simular procesos que permitan prever eventos.

2.3.2- PERFIL DEL ALUMNO

2.3.2.1- INGRESO

Los aspirantes a ingresar a la carrera de Ingeniería Matemática tienen mejores expectativas de un óptimo desempeño profesional si poseen las siguientes cualidades:

Gusto por las aplicaciones de las matemáticas.

Capacidad de concentración y razonamiento.

Creatividad, capacidad de análisis y síntesis.

Capacidad de trabajo en equipo o en forma interdisciplinaria.

Buena memoria y perseverancia.

Disciplina y constancia en el trabajo.

2.3.3- PERFIL DE EGRESO

2.3.3.1- CONOCIMIENTOS

Sólida formación profesional en matemáticas y conocimientos en el área según la opción que puede ser financiera o industrial.

Analizar problemas, crear modelos matemáticos que los representen, diseñar métodos para su solución, interpretar los resultados para plantear soluciones al problema.

Utilizar herramientas computacionales para resolver problemas matemáticos, así como desarrollar sus propios sistemas y aplicarlos a problemas específicos.

Dispondrá de elementos para juzgar el impacto, implicaciones y consecuencias de la aplicación de sus conocimientos y soluciones a problemas.

2.3.3.2- HABILIDADES

Dirigir o participar en proyectos de investigación científica de punta.

Tendrá flexibilidad para adaptarse al mercado, habilidad negociadora y sentido práctico para implantar los esquemas que propone con el fin de realizar con éxito cada proyecto donde interviene.

Tendrá capacidad de comunicación y expresión con poder de convencimiento para manejar el diálogo y la concertación con otros profesionales dirigiendo o integrando grupos interdisciplinarios.

2.3.3.3- ACTITUDES Y VALORES

Abierta y amplio criterio ante nuevos problemas y situaciones mostrándose dispuesto al cambio para el progreso y bienestar, tanto social como personal.

Tendrá voluntad y metodologías para difundir la cultura matemática a no especialistas.

Practicará la ética profesional, juzgará los límites del derecho en lo que respecta al manejo y conocimiento de información y tomará las decisiones adecuadas.

Tendrá iniciativa y voluntad de búsqueda de nuevas soluciones a problemas.

UNIDAD III

FUNDAMENTOS TEORICOS

En la actualidad, las variables que se evalúan para la selección del personal, tiene mucho que ver con el Perfil del Puesto, pero ¿que es el perfil del puesto?

3.1- PERFIL DE PUESTO

Es el proceso sistemático mediante el cual se obtiene toda la información referida a un puesto de trabajo y que pueda ser relevante tanto en lo que se refiere a un óptimo rendimiento en el mismo, como para la seguridad, satisfacción y comodidad de sus ocupantes.

La estrategia para definir el perfil del puesto es la siguiente:

1- *Análisis del cargo*: inventario de los aspectos intrínsecos (contenido del cargo) y extrínsecos (requisitos que debe cumplir el aspirante al cargo, factores de especificaciones) del cargo. Lo importante para la selección es la información con respecto a los requisitos y las características que debe poseer el aspirante al cargo.

El proceso de selección se concentrará en la búsqueda y la evaluación de esas exigencias y en las características de los candidatos que se presenten.

2- *Aplicación de la técnica de los incidentes críticos*: esta técnica busca identificar las características deseables (que mejoran el desempeño) y las no deseables (que empeoran el desempeño) en los candidatos. Naturalmente, presenta el inconveniente de fundamentarse en el arbitrio del jefe inmediato.

3- *Análisis de la solicitud de empleado*: consiste en la verificación de los datos consignados en la solicitud, a cargo del jefe inmediato, especificando los requisitos y las características que el aspirante al cargo debe poseer.

4- *Análisis del cargo en el mercado*: consiste en examinar en otras compañías los contenidos, los requisitos y las características de un cargo que va a crearse en la empresa, del cual, como es lógico, no se tiene una definición a priori.

5- *Hipótesis de trabajo*: en caso de que ninguna de las alternativas anteriores pueda aplicarse, solo queda por emplearse la hipótesis de trabajo, es decir, una predicción aproximada del contenido del cargo y su exigibilidad con relación al ocupante, como simulación inicial.

La información que el organismo recibe respecto de los cargos y de sus ocupantes se transforma en una ficha de especificaciones del cargo o ficha profesiográfica, que debe contener las características psicológicas y físicas necesarias para que el aspirante pueda desempeñarse satisfactoriamente en el cargo considerado.

3.2- ELECCIÓN DE LAS TÉCNICAS DE SELECCIÓN:

Una vez obtenida la información acerca del cargo y del aspirante, y elaborada la ficha profesiográfica, el paso que sigue es la elección de las técnicas de selección más adecuadas al caso o la situación.

Presentaremos las técnicas de selección en cinco grupos bien definidos:

3.2.1 Entrevista de selección

Pese a ser subjetiva, la entrevista personal es el factor que más influye en la decisión final respecto de la aceptación o no de un candidato al empleo. La entrevista de selección debe ser dirigida con gran habilidad y tacto, para que realmente pueda producir los resultados esperados. Además, tiene innumerables aplicaciones: elección en el reclutamiento, selección, consejería y orientación, evaluación del desempeño, desvinculación, etc.

La entrevista es un sistema de comunicación ligado a otros sistemas en función de cinco elementos básicos:

1- La fuente: el candidato, el cual posee características de personalidad, limitaciones, hábitos, maneras de expresarse, historia, problemas, etc. En este elemento se origina el mensaje.

2- El transmisor: el instrumento de codificación que transforma el mensaje en palabras, gestos o expresiones. La capacidad verbal y de expresión del candidato y del entrevistador se relaciona con el modo de codificar la información para transmitirla.

3- *El canal*: en la entrevista hay al menos dos canales, las palabras (lo mas importante) y los gestos.

4- *El instrumento para descifrar*: los receptores de la información (entrevistado y entrevistador) pueden interpretar (descifrar) los mensajes de manera diferente.

5- *El destino*: a quien se pretende transmitir el mensaje.

En el transcurso de la entrevista los cinco pasos anteriores operan de manera probabilística y no determinista, ya que puede ocurrir un ruido o transferencia en uno o en todos.

Entrevistar es probablemente el método que más se utiliza en la selección, no obstante que en este proceso existe demasiada subjetividad. Los objetivos intangibles, atribuidos a la entrevista son importantes para el buen desempeño en el cargo, y la evaluación que una persona capacitada haga es mejor que ninguna.

Etapas de la entrevista de selección:

a) *Preparación*: Las entrevistas de selección, determinadas o no, deben ser preparadas o planeadas de alguna manera. Aunque el grado de preparación varíe, debe ser suficiente para determinar:

- Los objetivos específicos de la entrevista.
- El método para alcanzar el objetivo de la entrevista.
- La mayor cantidad posible de información acerca del candidato entrevistado.

b) Ambiente: puede ser de dos tipos.

- Físico: el local de la entrevista debe ser confortable y solo para ese fin.
- Psicológico: el clima de la entrevista debe ser ameno y cordial.

c) Desarrollo de la entrevista: es la etapa fundamental del proceso, en el que se obtiene la información que ambos componentes desean.

- La entrevista es dirigida: cuando el entrevistador sigue un desarrollo establecido previamente, el cual sirve como lista de verificación, y por lo general, utiliza un formulario que sigue el orden de la solicitud de empleo. Este proceso es sencillo, fácil y rápido porque exige hacer anotaciones mínimas, lo cual permite que el entrevistador se concentre en el sujeto y no en las anotaciones ni en la secuencia de la entrevista.

- La entrevista es libre o no dirigida: cuando sigue el curso de las preguntas-respuestas-preguntas, es decir cuando no hay un derrotero preestablecido para cada entrevista. En una entrevista inicialmente se busca establecer contacto con el candidato para obtener información respecto de su vida y de su carrera profesional. Su infancia, su educación, los grados obtenidos en las escuelas donde estudio, la manera como enfrento los trabajos que le fueron encomendados, las razones por las cuales se desvinculo de las empresas donde trabajo. No debe darse oportunidad para que períodos oscuros de su vida pasen inadvertidos. El papel del entrevistador es de importancia capital en la entrevista.

d) Terminación de la entrevista: Para su terminación existe una necesidad semejante.

- El entrevistador debe hacer una señal clara que indique el final de la entrevista. • El candidato tiene que recibir algún tipo de información referente a lo que debe hacer en el futuro.

e) *Evaluación del candidato*: A partir del momento en que el entrevistado se marcha, el entrevistador debe iniciar la tarea de evaluación del candidato, puesto que los detalles están frescos en su memoria.

3.2.2 Pruebas de conocimiento o de capacidad

Las pruebas de conocimiento tienen como objetivo evaluar el grado de nociones, conocimientos y habilidades adquiridos mediante el estudio, la práctica o el ejercicio.

1- Según la manera como las pruebas se apliquen, pueden ser:

- Orales: preguntas y respuestas verbales.
- Escritas: preguntas y respuestas escritas.
- De realización: ejecución de un trabajo, prueba de mecanografía, de taquigrafía, de diseño, o de manejo de un vehículo o fabricación de piezas.

2- En cuanto al área de conocimientos, las pruebas pueden ser generales cuando tienen que ver con nociones de cultura o conocimiento generales y específicas cuando indagan conocimientos técnicos directamente relacionados con el cargo en referencia.

3- En cuanto a la manera como se elaboran:

- Tradicionales: disertativo, expositivo.
- Objetivas: mediante pruebas objetivas.
- Mixtas.

3.2.3 Pruebas psicométricas:

Se basan en el análisis de muestras del comportamiento humano, sometiéndose a examen bajo condiciones normativas, verificando la aptitud, para intentar generalizar y prever como se manifestara ese comportamiento en determinada forma de trabajo.

La prueba psicométrica es una medida objetiva y estandarizada de una muestra de comportamiento.

Se basan en las diferencias individuales que pueden ser físicas, intelectuales y de personalidad, y analizan cómo y cuánto varía la aptitud del individuo con relación al conjunto de individuos, tomado como patrón de comparación.

Aptitud es la potencialidad o predisposición de la persona para desarrollar una habilidad o un comportamiento. La aptitud es innata.

Aptitud: • predisposición natural para determinado trabajo o tarea, • existe sin ejercicio, • se evalúa por medio de comparaciones, • permite pronosticar el futuro del candidato, • con el ejercicio se transforma en capacidad, • es el estado latente y potencial del comportamiento.

La aptitud debidamente ejercida a través de la práctica se transforma en capacidad.

Capacidad es la habilidad real de la persona en determinada actividad o comportamiento y se adquiere a partir del desarrollo de una aptitud mediante el entrenamiento o la practica.

Capacidad: • habilidad para realizar determinado trabajo, • surge del entrenamiento o del aprendizaje, • se evalúa mediante el rendimiento en el trabajo, • se refiere a la habilidad real del individuo, • es el resultado de ejercitar la aptitud, • es el estado real del comportamiento.

Una prueba de conocimiento ofrece un diagnóstico real de las habilidades del candidato, en tanto que una prueba de aptitud proporciona un pronóstico futuro de su potencial de desarrollo.

EJEMPLO:

Una prueba muy conocida es la BETA II-R. Es un instrumento psicométrico diseñado para medir las capacidades intelectuales generales de las personas relativamente analfabetas o de personas con dificultades con el idioma en el país. Puede también ser utilizado como una medida no verbal en poblaciones generales. Este instrumento proporciona una estimación global de la capacidad intelectual, expresada ya sea mediante un CI (Coeficiente Intelectual) o mediante un percentil.

El BETA II –R Consta de 6 tareas, la duración de la aplicación es de aproximadamente 30 min. No se cuenta con un nombre específico para cada tarea,. Sin embargo, cada tarea pretende medir diferentes aspectos de la capacidad intelectual no verbal del individuo y puede ser identificada por el tipo de problema:

- Laberintos
- Claves
- Figuras Geométricas
- Figuras incompletas
- Pares iguales y pares desiguales
- Objetos equivocados

Laberintos: Mide la capacidad de atención, planeación y previsión, tales como atención a instrucciones. Coordinación Viso-motriz combinada con previsión.

El tiempo destinado para esta tarea es de un minuto con 30 segundos.

Claves: Mide la capacidad para aprender con rapidez asociaciones de símbolos y formas. Permite medir también destreza viso-motriz, especialmente en trabajo de lápiz como puede ser trabajo de oficina. Mide la expectativa que se puede tener para un trabajo con rapidez y precisión en asociaciones y ejecución de símbolos, principalmente nuevos.

Para esta segunda tarea el tiempo es de dos minutos.

Figuras Geométricas: Es un criterio fuerte para evaluación de pensamiento abstracto. Explora la habilidad para la percepción, la capacidad de análisis, y la capacidad de síntesis. Intervienen aspectos lógicos, y de razonamiento aplicados a las relaciones especiales así como la formación de conceptos no verbales. Se puede considerar la parte más poderosa de la prueba sobre la inteligencia general.

Cuatro minutos es el tiempo destinado para la tercera tarea.

Figuras incompletas: Mide la capacidad de captar visualmente los objetos familiares, y para determinar la ausencia de detalles esenciales frente a los no esenciales o indiferentes. Depende en gran medida de la capacidad de atención y la concentración que puede tener una persona.

Dos minutos con treinta segundos es el tiempo para la tarea de figuras geométricas.

Pares iguales: Mide la habilidad para establecer relaciones clasificatorias para ajustes en los acontecimientos ambientales. Calificaciones altas no muestran capacidad de comprensión, capacidad de pensamiento asociativo, y capacidad para discriminar objetos distintos y también la percepción visual.

Para la cuarta tarea, el tiempo es de dos minutos.

Objetos equivocados: Mide las habilidades de percepción de cómo funcionan en la realidad los objetos concretos. Así como reaccionar a los principios físicos y mecánicos que rigen las situaciones diarias.

Para los objetos equivocados el tiempo es de tres minutos.

El tiempo para cada tarea debe ser tomado con mucha precisión, con el fin de que se observe con mucha precisión los límites de tiempo. El examinador debe contar con un cronometro o con un reloj con segundero exacto.

3.2.4- Pruebas de personalidad:

Las pruebas de personalidad pretenden analizar los diversos rasgos determinados por el carácter (rasgos adquiridos) y por el temperamento (rasgos innatos). Se denominan genéricas cuando revelan los rasgos generales de personalidad en una síntesis global; y específicas, cuando investigan determinados rasgos o aspectos de la personalidad como equilibrio emocional, interés, frustraciones, ansiedad, agresividad, nivel de motivación, etc.

Ejemplo:

El encargado de recursos humanos puede pedir que se realice un dibujo de una persona y que escriba un relato referente a la persona para poder determinar los rasgos del aspirante.

3.2.5- Técnicas de simulación:

Junto con los resultados de las pruebas psicológicas y de las entrevistas, el aspirante es sometido a una situación de simulación de algún acontecimiento generalmente relacionado con el futuro papel que desempeñara en la empresa, suministrando una expectativa más realista acerca de su comportamiento futuro en el cargo.

3.3- ANALISIS MULTIVARIADO

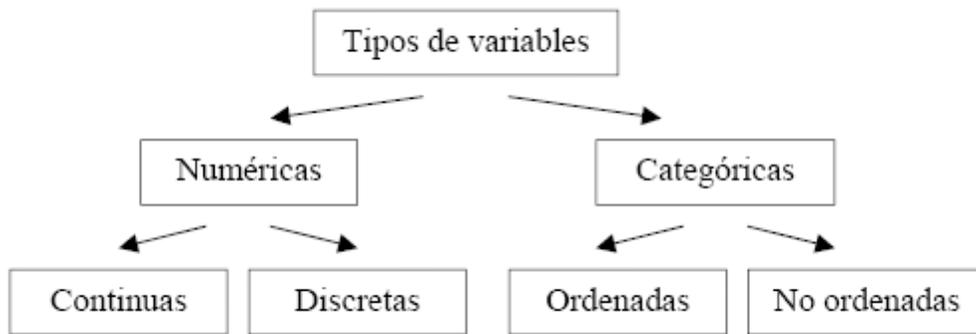
Los datos multivariados surgen en distintas áreas o ramas de la ciencia. Ejemplos:

- 1) Investigación de mercados: Identificar características de los individuos para determinar qué tipo de personas compran determinado producto.
 - 2) Agricultura: Resistencia de determinado tipo de cosechas a daños por plagas y sequías.
 - 3) Psicología: Relación entre el comportamiento de adolescentes y actitudes de los padres.
- ¿En qué situaciones surgen los datos multivariados?
 - Cuando a un mismo individuo se le mide más de una característica de interés.

 - Un individuo puede ser un objeto o concepto que se puede medir. Más generalmente, los individuos son llamados unidades experimentales. Ejemplos de objetos: personas, animales, terrenos, compañías, países, etc.
 - Ejemplos de conceptos: amor, amistad, noviazgo, etc.

 - Características de los individuos: Los individuos deben de ser independientes entre sí.

 - Una variable es una característica o atributo que se le mide a un individuo.



3.3.1-OBJETIVOS DE LOS MÉTODOS MULTIVARIADOS:

- 1) Simplificación: Los métodos multivariados son un conjunto de técnicas que permiten al investigador interpretar y visualizar conjuntos grandes de datos (tanto en individuos como en variables).
- 2) Relación: Encontrar relaciones entre variables, entre individuos y entre ambos.
 - 2.1) Relación entre variables: Existe relación entre variables cuando las variables miden una característica común. Ejemplo: Suponga que se realizan exámenes de lectura, ortografía, aritmética y álgebra a estudiantes de 6^o de primaria. Si cada uno de los estudiantes obtiene calificaciones altas, regulares o bajas en los cuatro exámenes, entonces los exámenes estarían relacionados entre sí. En este caso, la característica común que estos exámenes pueden estar midiendo podría ser la "inteligencia global".
 - 2.2) Relación entre individuos: Existe relación entre individuos si alguno de ellos son semejantes entre sí. Ejemplo: Suponga que se evalúan cereales (para el desayuno) respecto a su contenido nutricional y se miden, por ejemplo, los gramos de grasa, proteínas, carbohidratos y sodio a cada uno de ellos. Se

podría esperar que los cereales de fibra estén relacionados entre sí, o que los cereales endulzados tengan cierta relación entre sí, además se podría esperar que ambos grupos fueran diferentes de uno a otro.

- Uso de los métodos multivariados: Minerías de datos

- Los métodos multivariados son realmente un conjunto de técnicas que en su gran mayoría tienen un carácter exploratorio y no tanto inferencial.

3.3.2- CLASIFICACIÓN DE LOS MÉTODOS MULTIVARIADOS:

- 1) Dirigidas o motivadas por las variables: se enfocan en las relaciones entre variables. Ejemplos: matrices de correlación, análisis de componentes principales, análisis de factores, análisis de regresión y análisis de correlación canónica.
- 2) Dirigidas o motivadas por los individuos: se enfocan en las relaciones entre individuos. Ejemplos: análisis discriminante, análisis de cúmulos y análisis multivariado de varianza.

3.3.3-NOTACIÓN DE MATRICES Y VECTORES:

p = número de variables

n = número de individuos

X_{ij} = j -ésima variable del i -ésimo individuo

x_{ij} = valor observado de la j -ésima variable del i -ésimo individuo

$i=1,\dots,n$ y $j=1,\dots,p$

3.3.3.1- Matriz de datos:

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}$$

x_{ij} = elemento en el i-ésimo renglón y j-ésima columna

Renglones = individuos

Columnas = variables

3.3.3.2- Vectores de datos:

Los renglones de la matriz de datos se pueden expresar como vectores de la siguiente forma: El i-ésimo renglón de X se escribe como

$$x_i' = (x_{i1}, x_{i2}, \dots, x_{ip})$$

Nota: Todos los vectores son vectores columna, i.e.,

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}$$

3.3.4-ESPERANZAS Y VARIANZAS DE VECTORES ALEATORIOS.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

3.3.4.1-Media:

$$\mu = E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

μ es un vector de medias de dimensión $p \times 1$.

3.3.4.2-Varianzas-Covarianzas:

$$\Sigma = \text{Var}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X}') = E\{(\mathbf{X} - \mu)(\mathbf{X} - \mu)'\}$$

Escribiendo el vector completo,

$$\begin{aligned} \Sigma &= E \left\{ \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{pmatrix} (X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p) \right\} \\ &= E \begin{pmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \dots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \dots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \dots & (X_p - \mu_p)^2 \end{pmatrix} \end{aligned}$$

Finalmente, los elementos de Σ se denotan como:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

Donde,

$$\sigma_{jj} = \text{Cov}(X_j, X_j) = \text{Var}(X_j) = E\{(X_j - \mu_j)^2\}, \text{ para } j=1,2,\dots,p, \text{ y}$$

$$\sigma_{kj} = \text{Cov}(X_k, X_j) = E\{(X_k - \mu_k)(X_j - \mu_j)\}, \text{ para } k \neq j=1,2,\dots,p$$

es una matriz de varianzas y covarianzas dimensión $p \times p$.

3.3.4.3-Correlaciones:

$$P = \text{Corr}(X) = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}$$

donde,

$$\rho_{kj} = \text{Corr}(X_k, X_j) = \frac{\sigma_{kj}}{\sqrt{\sigma_{kk}} \sqrt{\sigma_{jj}}}, \text{ para } k \neq j=1,2,\dots,p$$

3.3.4.4-Comentarios:

- 1) El coeficiente de correlación ρ_{kj} es una medida de la relación lineal entre las variables X_k y X_j .
- 2) $-1 \leq \rho_{kj} \leq 1$
- 3) Si X_k y X_j son v.a. independientes $\rightarrow \rho_{kj} = 0$
- 4) $\rho_{kj} = 0 \rightarrow$ Independencia entre X_k y X_j únicamente en el caso Normal.
- 5) Para apreciar la relación (en general) entre dos variables es recomendable, además de calcular el coeficiente de correlación, hacer una gráfica de dispersión de ellas.

3.4- ANÁLISIS EXPLORATORIO MULTIVARIADO

3.4.1- Estadísticas multivariadas descriptivas

Las estadísticas descriptivas (multivariadas), como su nombre lo indica, sirven para describir el comportamiento de un conjunto de datos.

Formalmente, un conjunto de datos es una realización de una muestra aleatoria X_1, X_2, \dots, X_n de una distribución multivariada. Es decir, para $i = 1, 2, \dots, n$,

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{ip} \end{pmatrix}.$$

En otras palabras, cada X_i es una variable aleatoria multivariada de dimensión p .

Por lo tanto, un conjunto de datos esta formado por n realizaciones de p variables aleatorias.

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}.$$

3.4.1.1-MEDIA MUESTRAL:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

que en realidad, escribiendo el vector completo, se puede expresar como:

$$\hat{\mu} = \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_p \end{pmatrix} = \frac{1}{n} \left\{ \begin{pmatrix} \mathbf{X}_{11} \\ \mathbf{X}_{12} \\ \vdots \\ \mathbf{X}_{1p} \end{pmatrix} + \dots + \begin{pmatrix} \mathbf{X}_{n1} \\ \mathbf{X}_{n2} \\ \vdots \\ \mathbf{X}_{np} \end{pmatrix} \right\}.$$

Esto implica que, para $j=1, \dots, p$

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}.$$

Propiedades:

$$E(\hat{\mu}) = \mu.$$

3.4.1.2-VARIANZA MUESTRAL:

$$\hat{\Sigma} = \frac{1}{n-1} \left\{ \sum_{i=1}^n (\mathbf{X}_i - \hat{\mu})(\mathbf{X}_i - \hat{\mu})' \right\},$$

cuyos elementos se denotan como:

$$\hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \dots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \dots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \dots & \hat{\sigma}_{pp} \end{pmatrix}$$

donde,

$$\hat{\sigma}_{jj} = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \hat{\mu}_j)^2, \text{ para } j=1,2,\dots,p, \text{ y}$$

$$\hat{\sigma}_{kj} = \frac{1}{n-1} \sum_{i=1}^n (X_{ik} - \hat{\mu}_k)(X_{ij} - \hat{\mu}_j), \text{ para } k \neq j=1,2,\dots,p.$$

Propiedades:

$$E(\hat{\Sigma}) = \Sigma.$$

3.4.1.3-CORRELACIÓN MUESTRAL:

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

Donde, $r_{kj} = \frac{\hat{\sigma}_{kj}}{\sqrt{\hat{\sigma}_{kk}} \sqrt{\hat{\sigma}_{jj}}}$ para $k \neq j = 1, 2, \dots, p$

Propiedades:

1) $-1 \leq r_{kj} \leq 1$

2) $E(R) \neq P.$

3.4.2- ANÁLISIS GRÁFICO DE DATOS MULTIVARIADOS

3.4.2.1-DIAGRAMAS DE DISPERSIÓN (bidimensional).

Este tipo de diagrama consiste en graficar simultáneamente en dos dimensiones diagramas de dispersión entre todas las posibles parejas de variables.

3.4.2.2-DIAGRAMAS DE DISPERSIÓN (tridimensional)

Este tipo de diagrama consiste en graficar en tres dimensiones tres variables simultáneamente.

3.4.3- ANÁLISIS DE COMPONENTES PRINCIPALES.

3.4.3.1- Breve repaso de matrices

Sea Σ una matriz cuadrada de $p \times p$ tal que

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}.$$

Se dice que una matriz es simétrica si $\sigma_{jk} = \sigma_{kj}$ para todo $j, k = 1, 2, \dots, p$.

Las matrices de varianzas-covarianzas siempre son simétricas.

- Traza de una matriz:

$$\text{tr}(\Sigma) = \sum_{j=1}^p \sigma_{jj}.$$

- Determinante de una matriz (cuadrada):

$$\det(\Sigma) = |\Sigma| = \sum_{j=1}^p \sigma_{1j} \Sigma_{1j},$$

Donde:

$$\Sigma_{1j} = (-1)^{1+j} |\Sigma^{1j}| \text{ y } \Sigma^{1j}$$

es la matriz obtenida a partir de Σ al eliminar su primer renglón y su j -ésima columna.

El determinante de una matriz de 1×1 es igual al valor del único elemento.

$$\text{Ej: Si } \Sigma = (\sigma_{11}) \text{ entonces } |\Sigma| = \sigma_{11}.$$

El determinante de una matriz de 2×2 se calcula como:

$$\text{Si } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \text{ entonces } |\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}.$$

Ejemplo numérico:

$$\text{Sea } \Sigma = \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}.$$

Entonces, $\text{tr}(\Sigma) = 6 + 3 = 9$, y $|\Sigma| = (6)^{1+1}|3| + 2(-1)^{1+2}|2| = 18 - 4 = 14$.

- *Eigenvalores y eigenvectores*: Los eigenvalores (o valores característicos) y los eigenvectores (o vectores característicos) son valores y vectores que caracterizan una matriz (cuadrada) y satisfacen

$$\Sigma w = \lambda w,$$

donde λ es un eigenvalor y w es un eigenvector.

Los eigenvalores se obtienen como solución a la ecuación:

$$|\Sigma - \lambda \mathbf{I}| = 0,$$

donde \mathbf{I} es la matriz identidad. Esta expresión toma la forma de una ecuación polinomial en λ de grado p :

$$c_1 \lambda^p + c_2 \lambda^{p-1} + \dots + c_p \lambda + c_{p+1} = 0.$$

Las raíces de esta ecuación son los eigenvalores de Σ . En general,

$$\lambda' = (\lambda_1, \lambda_2, \dots, \lambda_p).$$

Si Σ es una matriz simétrica, sus eigenvalores son número reales y por lo tanto se pueden ordenar de forma descendente $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

Para cada eigenvalor λ_j , existe un eigenvector w_j que satisface la ecuación

➤ Propiedades:

$$\text{tr}(\Sigma) = \sum_{j=1}^p \lambda_j,$$

$$|\Sigma| = \prod_{j=1}^p \lambda_j.$$

Ejemplo numérico:

$$\text{Sea } \Sigma = \begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix}.$$

Los eigenvalores de Σ deben satisfacer $|\Sigma - \lambda I| = 0$, es decir,

$$\left| \begin{pmatrix} 6 - \lambda & 2 \\ 2 & 3 - \lambda \end{pmatrix} \right| = 0.$$

Esto implica que $(6 - \lambda)(3 - \lambda) - 4 = 0$, por lo que $\lambda^2 - 9\lambda + 14 = 0$. Resolviendo la ecuación obtenemos que $\lambda_1 = 7$ y $\lambda_2 = 2$.

Para calcular el eigenvector correspondiente a $\lambda_1 = 7$ hacemos, $\Sigma w_1 = \lambda w_1$, es decir,

$$\begin{pmatrix} 6 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{21} \end{pmatrix} = 7 \begin{pmatrix} w_{11} \\ w_{21} \end{pmatrix} \Rightarrow \begin{matrix} 6w_{11} + 2w_{21} = 7w_{11} \\ 2w_{11} + 3w_{21} = 7w_{21} \end{matrix} \Rightarrow w_{11} = 2w_{21}.$$

Existen muchos vectores que satisfacen la condición, pero el único vector normalizado

($w_1' w_1 = 1$) es: $w_1' = (2/\sqrt{5}, 1/\sqrt{5})$

Similarmente, resolviendo $\Sigma w_2 = \lambda_2 w_2$ para $\lambda_2 = 2$ se puede demostrar que

$w_2' = (1/\sqrt{5}, -2/\sqrt{5})$

- Una matriz es definida positiva si todos sus eigenvalores son positivos.
- Una matriz es semi-definida positiva si todos sus eigenvalores son no negativos.

NOTA: Las matrices de varianzas-covarianzas y de correlaciones tanto poblacionales como muestrales son semidefinidas positivas.

3.4.4- COMPONENTES PRINCIPALES

El análisis de componentes principales es un procedimiento matemático que transforma un conjunto de variables posiblemente correlacionadas en un conjunto menor de variables no correlacionadas llamadas componentes principales.

Dadas n observaciones de p variables, el objetivo del análisis de componentes principales es determinar r nuevas variables no correlacionadas llamadas componentes principales que representen la mayor variabilidad posible de las variables originales.

El uso de esta técnica es principalmente exploratoria y en general como un paso intermedio para análisis posteriores.

- **3.4.4.1- Los OBJETIVOS principales son:**
 - 1) Reducir la dimensionalidad de un conjunto de datos,
 - 2) Interpretar un conjunto de datos.

- **3.4.4.2- CARACTERÍSTICAS:**

- Las nuevas variables (componentes principales) son creadas de tal manera que:

- 1) No estén correlacionadas.

- 2) La 1^a componente principal explique la mayor variabilidad posible de los datos.

- 3) Cada componente subsecuente explique la mayor variabilidad posible restante no explicada por las componentes anteriores.

Formalmente, sea un vector aleatorio de p variables con matriz de varianzas-covarianzas Σ con eigenvalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Sean $Y' = (Y_1, Y_2, \dots, Y_p)$ nuevas variables formadas como combinaciones lineales de las X_i 's, i.e.,

$$\begin{aligned} Y_1 &= a_1'X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= a_2'X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= a_p'X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned}$$

Las componentes principales son aquellas combinaciones lineales Y_1, Y_2, \dots, Y_p no correlacionadas, cuyas varianzas son tan grandes como sea posible.

- **3.4.4.3- COMPONENTES:**

- 1^a componente principal: $Y_1 = a_1'X$, donde a_1 maximiza $Var(a_1'X)$ sujeto a

$$a_1'a_1 = 1$$

- 2^a componente principal: $Y_2 = (a_2'X)$, donde a_2 maximiza $Var(a_2'X)$ sujeto a

$$a_2'a_2 = 1 \text{ y } Cov(a_1'X, a_2'X) = 0$$

^a
 k componente principal: $Y_k = a_k' X$, donde a_k maximiza $Var(a_k' X)$ sujeto a

$$a_k' a_k = 1 \text{ y } Cov(a_k' X, a_j' X) = 0 \text{ para } j < k$$

Se puede demostrar que el máximo de la varianza de $a_1' X$ entre todos los vectores a_1 que satisfacen $a_1' a_1 = 1$ es igual a λ_1 y por lo tanto, a_1 es el eigenvector de Σ correspondiente al eigenvalor λ_1 .

También, se puede demostrar que el valor máximo de la varianza de $a_2' X$ entre todas las combinaciones lineales que satisfacen $a_2' a_2 = 1$ y que no están correlacionadas con Y_1 es igual a λ_2 . Por lo tanto, a_2 es el eigenvector de Σ correspondiente al eigenvalor λ_2 .

En general, se puede demostrar que el valor máximo de la varianza de $a_k' X$ entre todas las combinaciones lineales que satisfacen $a_k' a_k = 1$ y que no están correlacionadas con Y_1, Y_2, \dots, Y_{k-1} es igual a λ_k . Por lo tanto, a_k es el eigenvector de Σ correspondiente al eigenvalor λ_k .

▪ **3.4.4.4- INTERPRETACIÓN de λ_k :**

Recuerde que $tr(\Sigma) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp}$ es una medida de la variabilidad total de las variables originales. Por otro lado, $Var(Y_k) = Var(a_k' X) = \lambda_k$, $k=1, \dots, p$. Por lo tanto, la variabilidad total de las variables componentes principales $tr(\Sigma) = \lambda_1 + \lambda_2 + \dots + \lambda_p$ es igual a la variabilidad total de las variables originales.

$$\square \left(\begin{array}{l} \text{Proporción de la variabilidad} \\ \text{total explicada por la } k \text{-ésima} \\ \text{componente principal} \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad k=1, 2, \dots, p$$

- INTERPRETACIÓN del vector de pesos $a'_k = (a_{k1}, a_{k2}, \dots, a_{kp})$:

Los elementos a_{kj} del eigenvector a_k son llamados pesos y miden la importancia de la j -ésima variable en el k -ésimo componente principal.

¿CUÁNTOS componentes principales son suficientes?

El número de componentes principales que de alguna manera pudieran reemplazar a las variables originales, sin mucha pérdida de información, depende del problema en particular. En general, se desea que el porcentaje de la variabilidad explicada por los r primeros componentes sea de al menos el 80%.

Una forma alternativa de decidir el número de componentes significativos es graficando λ_k vs. k . Cuando los puntos de la gráfica tienden a nivelarse, estos eigenvalores suelen estar suficientemente cercanos a cero como para que puedan ignorarse.

NOTA: Si no se tiene la matriz de varianzas-covarianzas poblacional Σ , se realiza todo el análisis anterior sobre la matriz de varianzas-covarianzas muestral $\hat{\Sigma}$. En este caso, los componentes obtenidos serían estimaciones de los componentes poblacionales.

- **3.4.4.5- VALORES O MARCADORES (scores) de los componentes principales**

-

- Para poder visualizar las componentes principales es necesario calcular el valor de cada componente para cada individuo en un conjunto de datos.

Sea x_i el vector de variables medidas para cada individuo. Entonces el valor de la k -ésima componente principal para el i -ésimo individuo es

$$y_{ik} = a_k' x_i, \text{ para } i=1, \dots, n \text{ y } k=1, \dots, p.$$

3.4.5- COMPONENTES PRINCIPALES SOBRE VARIABLES ESTANDARIZADAS

Si la escala en que están medidas las variables no es uniforme (similar), es recomendable realizar un análisis de componentes principales sobre las variables estandarizadas, i.e.,

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}, Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}, \dots, Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

En notación matricial, $Z = \Sigma^{-1/2}(X - \mu)$.

Propiedades: $E(Z)=0$ y $\text{Var}(Z)=\text{Cov}(Z)=P$, donde P es la matriz de correlaciones de los datos originales X .

Los componentes principales $Y^* = (Y_1^*, Y_2^*, \dots, Y_p^*)$ del conjunto de variables estandarizadas $Z' = (Z_1, Z_2, \dots, Z_p)$ se obtienen de los eigenvectores de la matriz de correlación P de X.

Vectores de correlaciones de componentes:

Si λ_k' y a_k^* son los eigenvalores y eigenvectores de la matriz P, las cantidades $C_k = \lambda_k'^{1/2} a_k^*$ dan las correlaciones entre las variables estandarizadas y la k-ésima componente principal, i.e.,

$$\text{Corr}(Y_k^*, Z_j) = c_{kj}, \text{ para } j, k = 1, \dots, p.$$

NOTA: Los componentes principales obtenidos a partir de la matriz Σ son, en general, diferentes a los obtenidos de la matriz P.

UNIDAD IV

4.1 DESARROLLO DEL PROBLEMA

Los datos para poder resolver el problema propuesto son proporcionados por la empresa EXEL SERVIGRAFICA S.A. DE C.V. Con dirección en Democracia 61-A, Col. San Miguel Amantla, Azcopotzalco México D. F., C. P. 02700.

El giro de la empresa Exel Servigrafica S. A. DE C. V. es el de la impresión de las artes graficas.

El desarrollo del problema se lleva acabo con base en la recolección de datos de las pruebas psicometricas aplicadas a candidatos para el puesto de ayudantes en el área de terminado. Dichas pruebas psicometricas fueron evaluadas con la BETA II-R y el resultado en cada una las distintas tareas es el que se representa en la tabla de datos que a continuación se presentara.

Se utilizara el método de las componentes principales para poder determinar al o a los mejores candidatos para poder ser contratados para el puesto.

Los resultados se presentan en la siguiente tabla.

Aspirante	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6
1	3	52	5	6	12	12
2	5	66	8	6	17	9
3	9	89	8	13	20	12
4	7	86	13	8	13	18
5	6	79	6	12	20	15
6	0	32	5	7	0	11
7	6	66	10	13	13	13
8	5	59	2	7	5	13
9	9	90	11	11	21	14
10	7	45	6	9	13	11
11	6	70	14	13	17	9
12	7	77	16	18	25	19
13	6	58	6	8	8	9
14	5	49	8	8	15	20
15	4	51	9	9	11	12
16	9	61	6	12	11	15
17	6	69	8	12	18	11
18	8	45	10	10	13	13
19	7	60	9	5	7	8
20	6	87	13	10	23	13
21	8	78	16	16	16	8
22	7	58	7	12	13	16
23	8	80	16	15	20	4
24	8	52	11	10	16	18
25	3	39	2	5	8	8
26	6	70	7	14	10	12

27	9	69	14	13	15	19
28	8	58	8	6	7	11
29	8	57	9	9	12	12
30	5	40	6	7	10	9

Tabla 1. Resultado de las 6 tareas presentadas en la prueba psicometrica BETA II-R aplicada a 30 candidatos para el lugar ofrecido por la empresa.

Donde:

Tarea 1 = Laberintos

Tarea 2 = Claves

Tarea 3 = Figuras Geométricas

Tarea 4 = Figuras incompletas

Tarea 5 = Pares iguales y pares desiguales

Tarea 6 = Objetos equivocados

La forma de evaluar las distintas tareas únicamente le compete al departamento de recursos humanos, por tanto no será descrita en el presente proyecto.

Aplicando el análisis de componentes principales a los datos que se tienen obtendremos a los mejores candidatos.

Con este método se lograra transformar al conjunto de variables que posiblemente esten relacionadas a un conjunto menor de variables no correlacionadas.

Dado que los datos se encuentran en una misma escala de medición, se aplicara directamente el análisis de componentes principales a la matriz de datos. En el caso de que no estuvieran en la misma escala de medición, se tendría que estandarizar dichos datos. Llamaremos X a la matriz de datos que representan las calificaciones de los aspirantes a obtener empleo en la empresa de Exel Sevigrafica.

Dicha matriz es la siguiente:

X=

3	52	5	6	12	12
5	66	8	6	17	9
9	89	8	13	20	12
7	86	13	8	13	18
6	79	6	12	20	15
0	32	5	7	0	11
6	66	10	13	13	13
5	59	2	7	5	13
9	90	11	11	21	14
7	45	6	9	13	11
6	70	14	13	17	9
7	77	16	18	25	19
6	58	6	8	8	9
5	49	8	8	15	20
4	51	9	9	11	12
9	61	6	12	11	15
6	69	8	12	18	11
8	45	10	10	13	13
7	60	9	5	7	8
6	87	13	10	23	13
8	78	16	16	16	8
7	58	7	12	13	16
8	80	16	15	20	4
8	52	11	10	16	18
3	39	2	5	8	8
6	70	7	14	10	12
9	69	14	13	15	19
8	58	8	6	7	11
8	57	9	9	12	12
5	40	6	7	10	9

No es el objetivo mostrar paso por paso el desarrollo del método aplicado a los datos. Vasta mencionar que ya se dio una brebe explicación el capítulo anterior de lo que es el análisis de componentes principales.

En esta parte se explicaran los resultados obtenidos de las operaciones que se realizaran en paquetes como MINITAB y EXCEL.

Lo primero que se quiere es obtener son las componentes principales de la matriz X. Para ello se necesita la matriz de varianza-covarianza para des pues obtener vectores propios que determinaran las componentes principales.

El calculo de la matriz de varianza-covarianza que llamaremos $\hat{\Sigma}$ se da de la siguiente forma:

$$\hat{\sigma}_{ii} = \hat{Var}(x_i) = \frac{1}{p-1} \sum_{k=1}^p (x_{ki} - \hat{\mu}_i)^2$$

$$\hat{\sigma}_{ij} = \hat{cov}(x_i, x_j) = \frac{1}{p-1} \sum_{k=1}^p (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

Donde p = numero de variables y x representa a los valores de las distintas variables que se están manejando.

A continuación se presenta la matriz de varianza covarianza:

$$\hat{\Sigma} = \begin{vmatrix} 160.809 & 12.375 & 148.313 & 39.257 & 31.143 & 54.936 & 9.947 \\ 12.375 & 4.2402 & 17.8713 & 4.0126 & 3.3977 & 5.5529 & 1.5816 \\ 148.313 & 17.8713 & 244.823 & 35.589 & 30.577 & 62.853 & 6.692 \\ 39.257 & 4.0126 & 35.589 & 14.723 & 8.3839 & 13.677 & 1.6713 \\ 31.143 & 3.3977 & 30.577 & 8.3839 & 11.4988 & 12.051 & 2.3839 \\ 54.936 & 5.5529 & 62.853 & 13.677 & 12.0506 & 31.344 & 5.0391 \\ 9.947 & 1.5816 & 6.692 & 1.6713 & 2.3839 & 5.0391 & 14.5333 \end{vmatrix}$$

También se necesitara el vector de medias de las variables en cuestión.

El calculo del vector de medias se da de la siguiente forma:

$$\mu_r = \frac{\sum_{i=1}^{30} x_i}{30}$$

El vector de medias es el siguiente:

	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6
$\mu =$	6.37	63.07	8.97	10.13	13.63	12.47

Recordemos que los eigenvalores (valores característicos) y los eigenvectores (vectores característicos) son valores y vectores que caracterizan una matriz cuadrada y satisfacen:

$$\Sigma w = \lambda w$$

Donde λ es un eigenvalor y w es un eigenvector.

Los eigenvalores se obtienen como solución a la ecuación:

$$|\Sigma - \lambda I| = 0$$

Donde I es la matriz identidad. Esta ecuación toma la forma de una ecuación polinomial en λ de grado 6.

$$c_1 \lambda^6 + c_2 \lambda^5 + c_3 \lambda^4 + c_4 \lambda^3 + c_5 \lambda^2 + c_6 \lambda = 0$$

Las raíces de esta ecuación son los eigenvalores de Σ . En general,

$$\lambda' = (\lambda_1, \lambda_2, \dots, \lambda_6)$$

Como Σ es la matriz de varianza-covarianza, es simétrica y los eigenvalores son números reales y por tanto se pueden ordenar de forma descendente $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_6 \geq 0$

Para definir a las componentes principales, se dice que:

Sean $Y' = (Y_1, Y_2, \dots, Y_6)$ nuevas variables formadas como combinaciones lineales de las X_i 's y entonces se tiene:

$$Y_1 = a_1' X = a_{11} X_1 + a_{12} X_2 + \dots + a_{16} X_6$$

$$Y_2 = a_2' X = a_{21} X_1 + a_{22} X_2 + \dots + a_{26} X_6$$

:

$$Y_6 = a_6' X = a_{61} X_1 + a_{62} X_2 + \dots + a_{66} X_6$$

Las componentes principales son aquellas combinaciones lineales Y_1, Y_2, \dots, Y_6 no correlacionadas, cuyas varianzas son tan grandes como sea posible.

1^{ra} Componente Principal: $Y_1 = a_1'X$ donde a_1 maximiza la $Var(a_1'X)$ sujeto a

$$a_1'a_1 = 1$$

2^{da} Componente Principal: $Y_2 = a_2'X$, donde a_2 maximiza $Var(a_2'X)$ sujeto a

$$a_2'a_2 = 1 \text{ y } Cov(a_1'X, a_2'X) = 0$$

:

6^{ta} Componente Principal: $Y_6 = a_6'X$, donde a_6 maximiza $Var(a_6'X)$ sujeto a

$$a_6'a_6 = 1 \text{ y } Cov(a_j'X, a_6'X) = 0 \text{ para } j < 6$$

A continuación se presentan las componentes principales:

Variable	PC1	PC2	PC3	PC4	PC5	PC6
Tarea1	-0.072	-0.093	0.002	0.174	-0.003	0.978
Tarea2	-0.942	0.305	0.132	0.02	0.003	-0.044
Tarea3	-0.149	-0.361	-0.377	0.593	0.576	-0.149
Tarea4	-0.128	-0.338	-0.211	0.388	-0.813	-0.113
Tarea5	-0.261	-0.644	-0.263	-0.666	0.053	0.039
Tarea6	-0.032	-0.489	0.852	0.15	0.069	-0.077

TABLA 2: Componentes principales para la matriz de datos X

Uno de los objetivos del análisis de las componentes principales es reducir el número de variables, se utilizaran solo algunas de estas componentes principales. Para saber cuantas y cuales componentes se utilizaran, se muestra a continuación el resultado del analisis de los eigenvalores de la matriz de covarianza que tambien se obtiene del software MINITAB.

	PC1	PC2	PC3	PC4	PC5	PC6
Eigenvalue	273.63	20.17	12.69	7.66	4.49	2.53
Proportion	0.852	0.063	0.04	0.024	0.014	0.008
Cumulative	0.852	0.915	0.954	0.978	0.992	1

TABLA 3: Aquí se presenta el eigenvalor, la proporción de cada componente así como el valor acumulado.

Lo que se observa en los datos anteriores es que la proporción de la primera componente principal es del 85.2% y el de la segunda del 6.3%. Con estas dos componentes se explica un 91.5% de la variabilidad de los datos.

Recordemos que lo que el método recomienda que se tenga un número de componentes que expliquen por lo menos un 80% de la variabilidad de los datos.

Por lo tanto se tomaran los dos primeros componentes para determinar las calificaciones que arrojan dichas componentes para poder decidir a las personas que pudieran ser contratadas.

Los componentes principales que se tomaran son:

Variable	PC1	PC2
Tarea1	-0.072	-0.093
Tarea2	-0.942	0.305
Tarea3	-0.149	-0.361
Tarea4	-0.128	-0.338
Tarea5	-0.261	-0.644
Tarea6	-0.032	-0.489

TABLA 4: Componentes principales que explican el 91.5% de los datos.

Para estimar las calificaciones se realiza de la siguiente manera:

$$\hat{y}_{rj} = \hat{a}_j (X_r - \hat{\mu})$$

Donde

\hat{y}_{rj} = Calificación estimada para la r-ésima variable en el j-ésimo componente principal

\hat{a}_j = peso del j-ésimo componente principal

X_r = valor r-ésimo del vector de datos.

$\hat{\mu}$ = Media estimada de la j-ésima variable

El vector de medias sera el que ya anterior mente se había calculado.

	Tarea 1	Tarea 2	Tarea 3	Tarea 4	Tarea 5	Tarea 6
$\mu =$	6.37	63.07	8.97	10.13	13.63	12.47

El calculo de las calificaciones se realizara en Excel.

A continuación se presenta la tabla de resultados de las calificaciones.

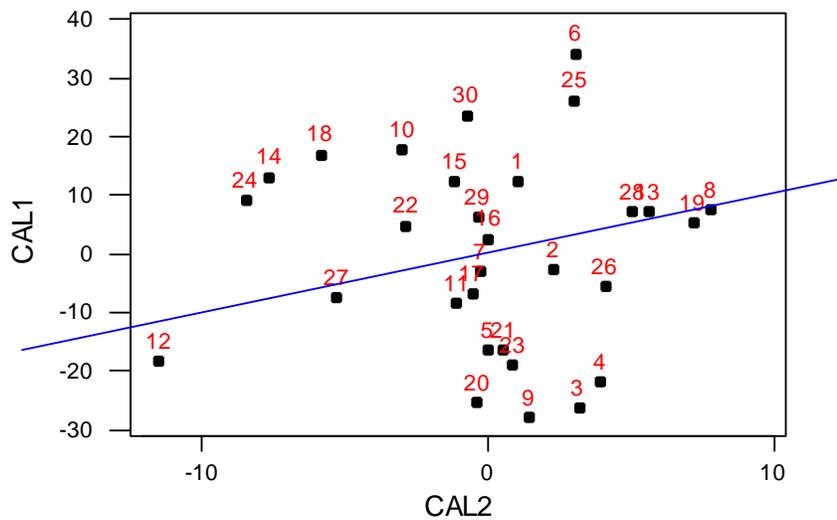
Aspirante	CAL1	CAL2
1	10.35	15.93
2	6.01	5.86
3	-4.44	-6.52
4	-0.56	4.55
5	1.23	0.97
6	16.79	27.33
7	2.64	5.52
8	8.55	19.62
9	-4.72	-5.80
10	6.76	16.23
11	-0.14	-4.42
12	-3.40	-5.76
13	6.35	12.94
14	9.19	21.90
15	8.80	14.45
16	2.24	13.90
17	2.93	2.21
18	5.17	16.29
19	5.41	12.95
20	-0.92	-6.68
21	-4.07	-9.41
22	4.51	14.03
23	-5.77	-16.63

24	4.06	16.17
25	13.17	21.19
26	2.99	6.06
27	-1.21	7.82
28	5.03	16.59
29	3.89	12.74
30	10.18	18.79

TABLA 5: Calificaciones de los 30 aspirantes estimadas a partir de las dos primeras componentes principales.

Para poder determinar al mejor o a los mejores, se representaran las calificaciones en una grafica.

La grafica de las calificaciones se realiza en MINITAB.



GRAFICA 1 : Calificación 1 VS Calificación 2

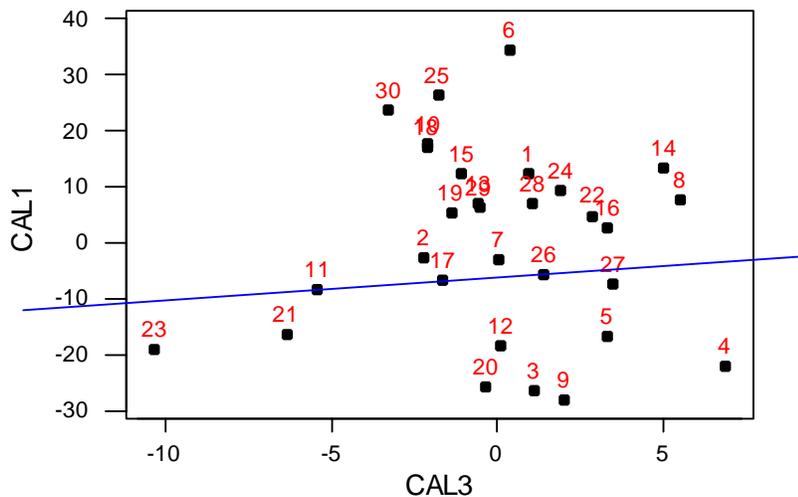
Para el análisis de la grafica, se realiza un trazo a 45° con respecto al origen, el mejor candidato para ocupar el puesto es el que este mas arriba y ala derecha sobre dicha recta. En algunos casos los mejores aspirantes se encuentran en el tercer cuadrante de la grafica de las calificaciones y el mejor es el que este mas abajo y a la izquierda como es el caso de las calificaciones presentadas anteriormente.

El porque primero se toma el que este mas arriba o mas abajo se debe a que sobre el eje vertical se coloco a la primera componente principal y recordemos que es la que mayor peso tiene.

En este caso el mejor candidato es el numero 12, y también se recomienda al aspirante 27.

Si se tomase en cuenta la componente principal PC3, se explicaría el 95.4% de la variabilidad de los datos.

A continuación se presentan las graficas con las 3 componentes principales:



GRAFICA 2 : Calificación 1 VS Calificación 3

En esta grafica, los aspirantes que se recomienda contratar son el 23, 21 y 11.

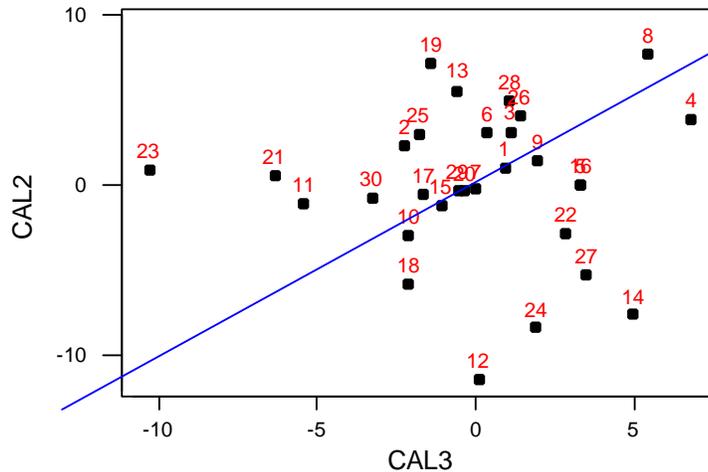


GRAFICO 3: Calificación 2 VS Calificación 3

En esta grafica no se encuentra un elemento que destaque mas que otros,. Esto se debe a que la suma de la proporción de las componentes 2 y 3 apenas alcanza el 10%.

CONCLUSIONES Y SUGERENCIAS:

Con al análisis de las componentes principales se pudo llegar a la recomendación para contratar a un aspirante a un puesto dentro de la empresa. El aspirante que se recomienda es el 12 y si se tuviera que contratar a algunos mas, se recomiendan las personas número 27, 23, 21 y 11 respectivamente.

Se puede aplicar al análisis de componentes principales para un mayor número de aspirantes y con un mayor número de variables, entiéndase por variable a los aspectos que se califiquen.

ANEXOS

<http://www.manpower.com/mpcom/files?name=GE-Brochure-Spanish.pdf>

<http://www.elprisma.com/apuntes/curso.asp?id=4371>

<http://www.cop.es/perfiles/contenido/trabajo.htm>

<http://www.perfilesplus.com/tour2.php>

<http://www.humsoftware.com/Paginas/Sel-Psicometricos.htm>

http://forteza.sis.ucm.es/Profes/speredadocencia_0203/trabajo/trabajo_practicas/practicas_trabajo.htm

Apuntes del seminario de titulación.