

PROTOCOLO PARA REALIZAR ANÁLISIS FACTORIAL EN VARIABLES QUE AFECTAN LAS CONDICIONES LABORALES

PROTOCOL TO PERFORM FACTOR ANALYSIS ON VARIABLES THAT AFFECT WORKING CONDITIONS

*Luz Elena Gutiérrez López**

DOI: <https://doi.org/10.18041/1909-2458/ingeniare.26.6564>

RESUMEN

Este trabajo describe los pasos que se deben realizar para aplicar la técnica multivariada de análisis factorial. Presenta el listado de estadísticos y criterios que se deben tener en cuenta para validar si la técnica de análisis factorial es apropiada para el caso de estudio. La población objetivo son los individuos en edad laboral de la ciudad de Montería, departamento de Córdoba, Colombia. La muestra trabajada fue de 642 personas. Los objetivos del estudio abarcan la construcción del protocolo, la descripción de las variables que influyen en las condiciones laborales y por último una propuesta de modelo de medición con el fin de relacionar los indicadores o variables observadas y las variables latentes o factores.

Palabras clave:

Análisis multivariado; Factores; Correlación; Rotación

ABSTRACT

This work describes the steps that must be carried out to apply the multivariate factor analysis technique. It presents the list of statistics and criteria that must be taken into account to validate whether the factor analysis technique is appropriate for the case study. The target population is individuals of working age from the city of Montería, Córdoba department, Colombia. The sample was 642 individuals. The goals of the study included the construction of the protocol, the description of the variables that influence working conditions and finally a proposal for a measurement model in order to relate the indexes or observed variables and the latent variables or factors.

Keywords:

Multivariate analysis; Factors; Correlation; Rotation

Cómo citar este artículo:

L. Gutiérrez López. Protocolo para realizar análisis factorial en variables que afectan las condiciones laborales. *Ingeniare*, Año 15, No. 26, Junio 2019. pp. 13 - 33.

* Ingeniera de sistemas. Departamento de Ingeniería de Sistemas y Computación, Universidad del Norte.
Correo: egluz@uninorte.edu.co

1. INTRODUCCIÓN

Hace una década no existía en el país el Ministerio de Trabajo y la información del Departamento Administrativo Nacional de Estadística (DANE), para el primer trimestre de 2010, reportaba una tasa de desempleo del 13% y daba cuenta de 18,6 millones de colombianos ocupados. El panorama a septiembre de 2019 presenta una tasa de desempleo del 10.2% y una tasa de informalidad que oscila entre en el 41.4% y el 72.1% en las principales ciudades del país [1].

Los temas laborales de un país nunca estarán completamente resueltos y siempre habrá asuntos pendientes, así como expectativas de mejoría por parte de las organizaciones sindicales [2]. A pesar de las iniciativas del Gobierno por generar empleo, la tasa de informalidad es alta y las condiciones del trabajador están influenciadas por múltiples variables como tipo de contrato, pago de prestaciones por parte del empleador, cantidad de horas laborales, pago de horas extras.

La Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad - MESEP fue creada en enero de 2009 mediante un convenio DANE-DNP con el objetivo de evaluar los factores que afectaron la comparabilidad de las cifras de mercado laboral y pobreza con el paso de la Encuesta Continua de Hogares - ECH a la Gran Encuesta Integrada de Hogares - GEIH, y realizar los empalmes correspondientes. En su segunda fase de trabajo, la MESEP se concentró en el diseño de la nueva metodología de medición de pobreza monetaria [3].

La base conceptual del cuestionario fue el método indirecto de medición de pobreza. En este método existen dos alternativas metodológicas para la definición del nivel mínimo de recursos a partir del cual se supone la satisfacción de las necesidades básicas: el método del Consumo Calórico y el método del Costo de las Necesidades Básicas.

Con el propósito de conocer el panorama de Colombia en función de condiciones laborales se revisó la encuesta realizada por el DANE para el año 2017. El instrumento tenía como objetivo general recoger información de variables que permitieran medir la medición de pobreza monetaria de los colombianos. La cobertura geográfica de la encuesta fue de 13 áreas metropolitanas tradicionales, 11 ciudades nuevas incluidas, cabeceras, resto urbano y los 24 departamentos.

El presente estudio se enfocó en la data de Personas. Este archivo contiene las variables de la Gran Encuesta Integrada de Hogares -GEIH 2017 a nivel de personas, necesarias para realizar los procedimientos de conformación, depuración, imputación y empalme del ingreso, para el cálculo de la pobreza monetaria, así como las variables construidas a través de dichos procedimientos. En el marco de la MESEP y con el fin de poder identificar la incidencia de las transferencias de las instituciones sobre la pobreza y la desigualdad, se decidió que el proceso de imputación de ingreso por transferencias (iof3) se realizará de forma separada para hogares (iof3h) e instituciones (iof3i), a partir de la publicación de la pobreza monetaria 2011-2012. Es decir, el archivo de ingreso tendrá por separado las transferencias de hogares y de instituciones (tanto observadas como imputadas), que antes se presentaban unidas. La Tabla 1 describe la ficha técnica del cuestionario empleado en el

presente estudio.

2. CONCEPTOS RELEVANTES

Para el presente estudio se utilizará la técnica multivariada Análisis Factorial. A continuación, se describen algunos conceptos importantes para la contextualización del desarrollo del caso de estudio.

Tabla 1. Ficha técnica estudio

Área temática	Sociedad
Sub área temática	Pobreza y condiciones de vida
Data File	Personas
Número de variables	137
Número de registros	762.755

Fuente: Elaboración del autor

2.1. Análisis Factorial

El análisis factorial es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Los grupos homogéneos se forman con las variables que correlacionan mucho entre sí y procurando, inicialmente, que unos grupos sean independientes de otros.

El Análisis Factorial es, por tanto, una técnica de reducción de la dimensionalidad de los datos. Su propósito último consiste en buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en los datos [4].

Hay dos tipos principales de análisis basados en el modelo de factor común: el análisis factorial exploratorio (EFA, por sus siglas en inglés) y el análisis factorial confirmatorio (CFA, por sus siglas en inglés). Tanto el EFA como el CFA pretenden reproducir las relaciones observadas entre un grupo de indicadores con un conjunto más pequeño de variables latentes, pero se diferencian fundamentalmente por el número y la naturaleza de las especificaciones a priori y restricciones hechas en el modelo de factores.

El análisis factorial exploratorio (EFA) es un enfoque impulsado o explorado por los datos, de tal manera que no se hacen especificaciones en relación con el número de factores latentes o al patrón de las relaciones entre los factores comunes y los indicadores (es decir, las cargas factoriales). Más bien, el investigador emplea el EFA como técnica exploratoria o descriptiva para determinar el número adecuado de factores comunes y descubrir cuáles variables de medición son indicadores razonables de las diversas dimensiones latentes.

En el análisis factorial confirmatorio (CFA), el investigador especifica el número de factores y el patrón de relación entre el indicador y las cargas factoriales de antemano, así como otros parámetros. El factor de solución preespecificado se evalúa en términos de lo bien que este reproduce la matriz de varianzas y covarianzas de la muestra en las variables medidas.

Por lo tanto, a diferencia del EFA, el CFA requiere una sólida base empírica o conceptual para guiar la especificación y evaluación del modelo de factores. En consecuencia, el EFA se utiliza en la etapa previa del proceso de desarrollo de la escala y construcción del constructo, mientras el CFA se utiliza en etapas posteriores, después de que la estructura subyacente ha sido previamente establecida a través bases teóricas y empíricas previas [5].

3. DESCRIPCIÓN DEL PROBLEMA

Teniendo en cuenta el panorama presentado en la introducción de este trabajo es importante analizar cuáles son los factores que influyen en las condiciones laborales de los colombianos y como se relacionan entre ellas. A continuación, se listan algunas de las preguntas que se pretenden responder en la presente investigación:

1. ¿Existe relación entre el salario y el nivel de estudio?
2. ¿Existe relación entre el estrato y el nivel de estudio?
3. ¿Existe relación entre el género y el nivel de estudio?

3.1 Objetivos de la Investigación

Con el fin de aportar en la solución del problema planteado se plantean 3 objetivos para la investigación:

1. Diseñar un protocolo para la identificación de factores que conecten las variables del presente estudio.
2. Identificar cuáles son las variables que intervienen en las condiciones laborales y cómo se relacionan entre ellas.
3. Construir el modelo de medida de las variables factoriales y observables del presente estudio.

4. PROCESAMIENTO DE ARCHIVOS

A continuación, se relacionan los pasos a seguir con el fin de generar los archivos de trabajo para esta investigación.

1. Debido a que las variables registradas en el archivo original del DANE solo contienen en la cabecera del documento los códigos de las variables, se deben asignar nombres explicativos a las variables. Por ejemplo, la variable P6020 en adelante se denominará Género.
2. Revisar calidad de las variables. Se encontraron 6 variables repetidas. P7472, P7472s1, P7500s1, P7510s2, P7510s2a1, P7510s7a1
3. Se eliminan las columnas de variables repetidas identificadas en el paso 2.
4. Se eliminaron las variables de manejo interno del DANE porque no aportaban información al proyecto: Cclasnr2, Cclasnr3, Cclasnr4, Cclasnr5, Cclasnr6, Cclasnr7, Cclasnr8, Cclasnr11, Impaes, Isaes, lees, Imdies, lof1es, lof2es, lof3hes, lof3ies y lof6es.
5. Se seleccionan las ciudades objeto del estudio, porque inicialmente se buscaba comparar entre ciudades.

No obstante, luego de revisar la cantidad de registros se focaliza el estudio hacia una sola ciudad.

6. Con el fin de utilizar una muestra representativa de la ciudad de Montería, se procede a calcular el número de registros con los que se va a trabajar. La Figura 1 presenta la herramienta utilizada para el cálculo de la muestra [6].

Tabla 2. Cantidad registros por ciudades del Caribe

No.	Ciudad	Cantidad de registros
1	Barranquilla	39.100
2	Cartagena	28.821
3	Montería	26.934
4	Riohacha	28.850
5	Santa Marta	30.665
6	Sincelejo	29.018
7	Valledupar	26.708

Fuente: Elaboración del autor

Figura 1. Tamaño muestra Montería

Calcula el tamaño de la muestra

Tamaño de la población: 26934

Nivel de confianza (%): 95

Margen de error (%): 3

Tamaño de la muestra: **1.027**

Fuente: Elaboración del autor

7. El paso siguiente es seleccionar aleatoriamente 1028 registros de la población de 26.934. Se utilizó el método propuesto en [7] para realizar la selección aleatoria de los registros. El método se describe en la Tabla 3.

8. Se debe definir la población objetivo que para este caso es la población en edad de trabajar. Es decir, se requieren los individuos en el rango de edad de 18 a 65 años. Luego de aplicar ese filtro quedan 642 registros.

9. En este punto se tienen 114 variables para analizar. Luego de una segunda revisión se elimina la variable: P550, ganancia de la cosecha anual. Porque no aporta al objeto del estudio. Por tanto, quedan 113 variables.

10. Luego se procede a revisar cada variable del archivo teniendo como base el objeto del presente estudio con el fin de identificar cuáles son las variables que pueden aportar información relevante para el análisis.

Tabla 3. Pasos selección aleatoria en Excel

Paso 1	Insertar columna al inicio del archivo.
Paso 2	Seleccionar la columna A
Paso 3	Escribir "= RAND ()" (sin comillas) en el cuadro de texto "Fórmula". Pulsar la tecla "Enter" para generar los números aleatorios en la columna A
Paso 4	Seleccionar todos los datos en la hoja de cálculo, junto con los números aleatorios correspondientes. No seleccionar los títulos ni los encabezados.
Paso 5	Utilizar la pestaña "Datos" en la parte superior de la pantalla y hacer clic en el botón "Ordenar" del grupo "Ordenar y filtrar" de la cinta "Datos". El cuadro de diálogo "Ordenar" se abrirá en la pantalla.
Paso 6	Seleccionar la opción "columna A" de la lista desplegable "Ordenar por columna" y "Del más pequeño al más grande" de la lista desplegable "Ordenar por Orden". Hacer clic en el botón "Aceptar" para cerrar el cuadro de diálogo y volver a la hoja de cálculo. Elegir el número máximo de filas que componen la muestra aleatoria.

Fuente: Elaboración del autor

La Tabla 4 presenta las variables de interés (22 variables), con las opciones de respuesta registradas en la encuesta nacional del DANE y su respectiva recodificación para ser usadas en el análisis de la investigación. La variable directorio no se utilizará para los procesos estadísticos, será una variable de control e identificación de los individuos.

Tabla 4. Descripción variables del estudio

ID DANE	Nombre Variable	Opciones respuesta encuesta	Recodificación
Directorio	Identificador del registro	Número generado por el DANE	No se recodifica
Estrato1	Estrato	1. estrato 1 2. estrato 2 3. estrato 3 4. estrato 4 5. estrato 5 6. estrato 6	No se recodifica
P6020	Género	1. Hombre 2. Mujer	No se recodifica
P6040	Edad	18 a 65 años	1. Joven (18 a 25 años) 2. Adulto (26 a 65 años)
P6090	SeguridadSocial	1 Sí 2 No 9 No sabe, no informa	1. Sí 2. No 3. No sabe, no informa (9, celdas vacías)
P6210	UltimoGradoEstudios	a. Ninguno b. Preescolar c. Básica primaria (1o - 5o) d. Básica secundaria (6o - 9o) e. Media (10o - 13o) f. Superior o universitaria g. No sabe, no informa – 0 para SPSS	1. No sabe, no informa (1, 7, celdas vacías) 2. Preescolar 3. Básica primaria 4. Secundaria (4 y 5) 5. Superior (6)

Tabla 4. Descripción variables del estudio (continuación)

ID DANE	Nombre Variable	Opciones respuesta encuesta	Recodificación
P6240	TipoActividad	a. Trabajando b. Buscando trabajo c. Estudiando d. Oficios del hogar e. Incapacitado permanente para trabajar f. Otra actividad	1. Trabajando 2. Buscando trabajo 3. Estudiando 4. Oficios del hogar 5. Incapacitado permanente para trabajar 6. Otra actividad 7. No sabe, no informa (celdas vacías)
P6426	TiempoTrabajo	De 0 a 504 meses	1. 0 – 60 2. 61 – 120 3. 121 – 180 4. 181 – 240 5. 241 – 300 6. 301 – 360 7. 361 – 420 8. 421 – 480 9. 481 - 600
P6430	TipoEmpleado	a. Obrero o empleado de empresa particular b. Obrero o empleado del gobierno c. Empleado doméstico d. Trabajador por cuenta propia e. Patrón o empleador f. Trabajador familiar sin remuneración g. Trabajador sin remuneración en empresas o negocios de otros hogares h. Jornalero o peón i. Otro	1. Obrero o empleado de empresa particular 2. Obrero o empleado del gobierno 3. Empleado doméstico 4. Trabajador por cuenta propia 5. Patrón o empleador 6. Trabajador familiar sin remuneración 7. Trabajador sin remuneración en empresas o negocios de otros hogares 8. Jornalero o peón

Tabla 4. Descripción variables del estudio (continuación)

ID DANE	Nombre Variable	Opciones respuesta encuesta	Recodificación
			9. Otro 10. no sabe, no informa (vacías)
P6545	Primas	1. Sí 2. No 9. No sabe, no informa	1. Sí 2. No 3. No sabe, no informa (vacías)
P6585s2	SubTransporte	1. Sí 2. No 9. No sabe, no informa	1. Sí 2. No 3. No sabe, no informa (vacías)
P6630s1	PrimaServicio	1. Sí 2. No	1. Sí 2. No 3. No sabe, no informa (vacías)
P6630s2	PrimaNavidad	1. Sí 2. No	1. Sí 2. No 3. No sabe, no informa (vacías)
P6630s3	PrimaVacas	1. Sí 2. No	1. Sí 2. No 3. no sabe, no informa (9, vacías)
P6800	HorasLaborales	5 a 72 horas/semana	1. No sabe, no informa (celdas vacías) 2. De 1 a 20 3. 21 – 48 4. 49 – 60 5. Mas de 60
P6920	Pensiones	1. Sí 2. No 3. Ya es pensionado	1. Sí 2. No 3. Ya es pensionado 4. No sabe, no informa (vacías)

Tabla 4. Descripción variables del estudio (continuación)

ID DANE	Nombre Variable	Opciones respuesta encuesta	Recodificación
P7040	OtraOcupación	1. Sí 2. No	1. Sí 2. No 3. No sabe, no informa (vacías)
P7150	CambioTrabajo	1. Sí 2. No	1. Sí 2. No 3. No sabe, no informa (vacías)
Oc	ocupado	1. Ocupado	1. Ocupado 2. No sabe, no informa (vacías)
Des	desocupado	1. Desocupado	1. Desocupado 2. No sabe, no informa (vacías)
Ingtot	IngresosTotal	0 a 20.000.000 millones	1. Menos de 1 SMMLV 2. De 1 a 3 SMMLV 3. De 3 a 5 SMMLV 4. De 5 a 7 SMMLV 5. De 7 a 10 SMMLV 6. Más de 10 SMMLV 7. No sabe/no informa (cero, vacías)

Fuente: Elaboración del autor

11. Se copian las variables de interés en el software estadístico SPSS para realizar el proceso de recodificación.

5. ANÁLISIS FACTORIAL EXPLORATORIO

La técnica de análisis factorial tiene varios indicadores que deben revisarse para decidir si la técnica es apropiada para la estudio a realizar. Entre estos criterios están: valor del alfa de Cronbach, valores de la matriz de correlación, comparación de chi cuadrado, KMO, análisis de comunalidad y varianza total explicada [8]. En las siguientes secciones se describirán cada uno de estos indicadores.

5.1 Análisis de Fiabilidad

Es importante garantizar la fiabilidad de los datos del estudio, para realizar esta validación se utiliza el estadístico denominado Alfa de Cronbach. Este estadístico supone un modelo de consistencia interna que estima el límite inferior del coeficiente de fiabilidad basándose en la varianza o en el promedio de las correlaciones entre los ítems, y cuyo cálculo viene determinado por la expresión de la parte izquierda de la Figura 2. En el presente estudio se calculó usando la herramienta SPSS en el menú Analizar => Escala => Análisis de Fiabilidad. Donde α representa el valor del alfa de Cronbach, n es el número de ítems y r el promedio de la correlación lineal de cada uno de los ítems.

Figura 2. Alfa de Cronbach

$$\alpha = \frac{nr}{1 + r(n-1)}$$

Alfa de Cronbach	N de elementos
.804	21

Fuente: Elaboración del autor

5.2 Estadísticos descriptivos

Se utilizó una muestra de 642 individuos. Todos en edad de trabajar, clasificados en joven (18 a 25 años) con un 24% y adulto (26 a 65 años) con 76%. Hombres 45% y mujeres 55%. El estrato de mayor incidencia fue el 1 con 67%, estrato 2 con 20%, estrato 3 con 9% y los estratos 4, 5 y 6 con un 1% cada uno. El nivel de formación de los encuestados en su mayoría era de secundaria con un 53%, básica primaria con 14%, formación superior de 29% y un 4% de datos perdidos. La Figura 3 presenta para cada variable el número de registros observados, el mínimo y máximo de los rangos, la media, desviación estándar y la varianza.

5.3 Análisis de la Matriz de Correlación

Los determinantes próximos a cero indican que las variables utilizadas están linealmente relacionadas, lo que significa que el análisis factorial, es una técnica pertinente para analizar esas variables. Para el caso de estudio el determinante es de 4.341E-14, siendo este un valor positivo para seguir con estas variables. Las correlaciones son diferentes de cero; esto también es un indicador positivo (Ver Figuras 4.a y 4.b).

La medida de adecuación muestral KMO (Kaiser-Meyer-Olkin) contrasta si las correlaciones parciales entre las variables son suficientemente pequeñas. Permite comparar la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial. EL estadístico KMO varía entre 0 y 1. Los valores pequeños indican que el análisis factorial puede no ser la técnica más adecuada, dado que las correlaciones entre los pares de variables no pueden ser explicadas por otras variables. Los menores que 0.5 indican que no debe utilizarse el análisis factorial con los datos muestrales que se están analizando. De acuerdo con la Figura 5 la significancia es excelente debido a que es cercana a 0 y el valor del KMO es excelente con un valor de 0.918. Por tanto, se puede seguir con el análisis.

Figura 3. Estadísticos descriptivos del estudio

	N	Mínimo	Máximo	Media	Desviación estándar	Varianza
Estrato	642	1	6	1,52	,927	,859
Genero	642	1	2	1,55	,498	,248
Edad	642	1,00	2,00	1,7586	,42829	,183
Seguridad	642	1,00	2,00	1,0405	,19728	,039
GradoEstudios	642	1,00	5,00	4,0234	,89255	,797
TipoActividad	642	1,00	6,00	2,1791	1,56123	2,437
TiempoTrabajo	642	1,00	10,00	4,3988	3,99374	15,950
TipoEmpleado	642	1,00	10,00	5,0841	3,58265	12,835
Primas	642	1,00	3,00	2,6449	,49495	,245
SubTransporte	642	1,00	3,00	2,5047	,73945	,547
PrimaServicio	642	1,00	3,00	2,4579	,79915	,639
PrimaNavidad	642	1,00	3,00	2,6137	,56164	,315
PrimaVacas	642	1,00	3,00	2,6121	,56473	,319
HorasLaborales	642	1,00	5,00	2,4470	1,08363	1,174
Pensiones	642	1,00	4,00	2,3941	1,17122	1,372
OtraOcupacion	642	1,00	3,00	2,3006	,49487	,245
HorasAdicionales	642	1,00	3,00	2,3037	,48979	,240
CambioTrabajo	642	1,00	3,00	2,8411	,46700	,218
ocupado	642	1,00	2,00	1,3178	,46597	,217
desocupado	642	1,00	2,00	1,9128	,28239	,080
IngresosTotal	642	1,00	7,00	3,0296	2,43942	5,951
N válido (por lista)	642					

Fuente: Elaboración del autor

Figura 4.a. Valores matriz correlación (primera sección)

	Estrato	Genero	Edad	Seguridad	GradoEstudios	TipoActividad	TiempoTrabajo	TipoEmpleado	Primas	SubTransporte
Correlación										
Estrato	1,000	,000	,035	-,021	,217	-,103	-,075	-,085	-,090	-,037
Genero	,000	1,000	,067	-,100	-,004	,301	,156	,206	,085	,080
Edad	,035	,067	1,000	-,032	-,214	-,150	-,262	-,266	-,081	-,107
Seguridad	-,021	-,100	-,032	1,000	-,041	,027	,039	,066	,084	,095
GradoEstudios	,217	-,004	-,214	-,041	1,000	-,131	,028	-,050	-,151	-,179
TipoActividad	-,103	,301	-,150	,027	-,131	1,000	,664	,703	,466	,442
TiempoTrabajo	-,075	,156	-,262	,039	,028	,664	1,000	,910	,496	,467
TipoEmpleado	-,085	,206	-,266	,066	-,050	,703	,910	1,000	,725	,702
Primas	-,090	,085	-,081	,084	-,151	,466	,496	,725	1,000	,904
SubTransporte	-,037	,080	-,107	,095	-,179	,442	,467	,702	,904	1,000
PrimaServicio	-,088	,080	-,123	,100	-,216	,441	,453	,705	,925	,942
PrimaNavidad	-,117	,097	-,090	,085	-,175	,426	,451	,688	,954	,842
PrimaVacas	-,115	,094	-,085	,085	-,171	,426	,453	,689	,946	,851
HorasLaborales	,089	-,257	,250	-,019	-,053	-,698	-,858	-,841	-,463	-,420
Pensiones	-,127	,190	-,314	,079	-,098	,689	,888	,951	,629	,642
OtraOcupacion	-,080	,170	-,319	,051	,037	,665	,926	,901	,456	,425
HorasAdicionales	-,088	,165	-,312	,050	,019	,641	,933	,903	,452	,420
CambioTrabajo	,013	,088	-,005	,002	-,017	,182	,262	,186	,005	-,011
ocupado	-,095	,194	-,319	,046	,027	,692	,958	,937	,490	,458
desocupado	,059	,009	-,135	-,077	-,097	-,180	-,434	-,425	-,222	-,207
IngresosTotal	,020	,105	-,331	,040	,141	,449	,693	,666	,306	,286

Fuente: Elaboración del autor

Figura 4.a. Valores matriz correlación (segunda sección)

PrimaServicio	PrimaNavidad	PrimaVacas	HorasLaborales	Pensiones	OtraOcupacion	HorasAdicionales	CambioTrabajo	ocupado	desocupado	IngresosTotal
-.088	-.117	-.115	.089	-.127	-.080	-.088	.013	-.095	.059	.020
.080	.097	.094	-.257	.190	.170	.165	.088	.194	.009	.105
-.123	-.090	-.085	.250	-.314	-.319	-.312	-.005	-.319	.135	-.331
.100	.085	.085	-.019	.079	.051	.050	.002	.046	-.077	.040
-.216	-.175	-.171	-.053	-.098	.037	.019	-.017	.027	-.097	.141
.441	.426	.426	-.698	.689	.665	.641	.182	.692	-.180	.449
.453	.451	.453	-.858	.888	.926	.933	.262	.958	-.434	.693
.705	.688	.689	-.841	.951	.901	.903	.186	.937	-.425	.666
.925	.954	.946	-.463	.629	.456	.452	.005	.490	-.222	.306
.942	.842	.851	-.420	.642	.425	.420	-.011	.458	-.207	.286
1,000	.916	.916	-.422	.665	.432	.425	-.039	.463	-.210	.264
.916	1,000	.988	-.444	.621	.435	.433	-.014	.470	-.213	.269
.916	.988	1,000	-.443	.621	.435	.432	-.015	.469	-.212	.270
-.422	-.444	-.443	1,000	-.836	-.856	-.847	-.242	-.897	.403	-.636
.665	.621	.621	-.836	1,000	.899	.901	.163	.936	-.424	.628
.432	.435	.435	-.856	.899	1,000	.942	.227	.965	-.437	.696
.425	.433	.432	-.847	.901	.942	1,000	.300	.971	-.440	.700
-.039	-.014	-.015	-.242	.163	.227	.300	1,000	.232	-.105	.207
.463	.470	.469	-.897	.936	.965	.971	.232	1,000	-.453	.712
-.210	-.213	-.212	.403	-.424	-.437	-.440	-.105	-.453	1,000	-.322
.264	.269	.270	-.636	.628	.696	.700	.207	.712	-.322	1,000

Fuente: Elaboración del autor

Figura 5. Valores KMO y prueba esfericidad

Medida Kaiser-Meyer-Olkin de adecuación de muestreo		.918
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	19481,298
	gl	210
	Sig.	.000

Fuente: Elaboración del autor

La prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad, en cuyo caso no existirían correlaciones significativas entre las variables y el modelo factorial no sería pertinente. Como se puede apreciar en la Figura 6 el chi estimado es mayor que el chi tabulado, siendo este un resultado positivo. El valor del chi estimado calculado con Excel coincide con el entregado por SPSS (ver Figura 5).

Figura 5. Valores KMO y prueba esfericidad

	A	B	C	D	E	F
1	Fórmula de recurrencia					
2	n	642				
3	p	21				
4	Determinante	4,341E-14				
5	significancia	0				
6	Grados de libertad	210	p(p-1)/2			
7						
8	Chi estimado	19481,32681		Conclusión:		
9	Chi Tabulado	244,808		Chi estimado mayor que Chi tabulado		

Fuente: Elaboración del autor

5.4 Análisis de Comunalidad

Los valores de la comunalidad deben ser iguales o superior a 0.5, debido a que valores superiores a 0.5 reflejan el grado de aporte de la variable. Para el presente estudio la mayoría de las variables tienen comunalidades iguales o superiores a 0.5. Las variables Edad, Seguridad, CambioTrabajo y desocupado tienen valores menores a 0.5; estas variables serán revisadas en el proceso de rotación. Para el detalle revisar la Figura 7.

Figura 7. Valores comunalidad

	Inicial	Extracción
Estrato	1,000	,529
Genero	1,000	,572
Edad	1,000	,413
Seguridad	1,000	,309
GradoEstudios	1,000	,642
TipoActividad	1,000	,643
TiempoTrabajo	1,000	,912
TipoEmpleado	1,000	,972
Primas	1,000	,946
SubTransporte	1,000	,882
PrimaServicio	1,000	,947
PrimaNavidad	1,000	,934
PrimaVacas	1,000	,934
HorasLaborales	1,000	,840
Pensiones	1,000	,926
OtraOcupacion	1,000	,923
HorasAdicionales	1,000	,931
CambioTrabajo	1,000	,236
ocupado	1,000	,971
desocupado	1,000	,343
IngresosTotal	1,000	,628

Fuente: Elaboración del autor

5.5 Varianza Total Explicada

Para el análisis de componentes si el valor de la columna Autovalores iniciales - Total da igual o superior a 1, se puede decir que es un componente candidato. Como se puede apreciar en la Figura 8 serían 4 componentes. Verificando la varianza acumulada sería del 73.492%. Este valor es un buen indicador debido a que explica en más del 70% el proceso.

En el gráfico de sedimentación se puede visualizar la gráfica de la magnitud de los autovalores. El corte en la tendencia descendente sirve de regla para la determinación del número de factores óptimo que deben estar presentes en la solución. Siempre se muestra la representación de los autovalores de la matriz de correlaciones (o de covarianzas) originales, independientemente del método de extracción seleccionado.

Como un soporte a la revisión de los autovalores se analizan también los puntos de la gráfica de sedimentación que tienen autovalores mayores o iguales a 1 en el eje Y. La cantidad de puntos que cumplan esa condición representan el número de posibles factores.

Figura 7. Valores comunalidad

Componente	Autovalores iniciales			Sumas de extracción de cargas al cuadrado			Sumas de rotación de cargas al cuadrado		
	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado	Total	% de varianza	% acumulado
1	9,981	47,527	47,527	9,981	47,527	47,527	7,528	35,846	35,846
2	2,960	14,095	61,622	2,960	14,095	61,622	5,325	25,356	61,202
3	1,337	6,366	67,988	1,337	6,366	67,988	1,336	6,364	67,566
4	1,156	5,504	73,492	1,156	5,504	73,492	1,244	5,926	73,492
5	,998	4,752	78,244						
6	,880	4,190	82,434						
7	,830	3,951	86,385						
8	,762	3,628	90,013						
9	,603	2,872	92,885						
10	,442	2,107	94,992						
11	,356	1,696	96,688						
12	,224	1,067	97,755						
13	,162	,772	98,526						
14	,092	,438	98,964						
15	,061	,289	99,252						
16	,051	,245	99,497						
17	,035	,165	99,662						
18	,032	,154	99,816						
19	,021	,102	99,918						
20	,010	,048	99,966						
21	,007	,034	100,000						

Fuente: Elaboración del autor

En el gráfico de sedimentación se puede visualizar la gráfica de la magnitud de los autovalores. El corte en la tendencia descendente sirve de regla para la determinación del número de factores óptimo que deben estar presentes en la solución. Siempre se muestra la representación de los autovalores de la matriz de correlaciones (o de covarianzas) originales, independientemente del método de extracción seleccionado.

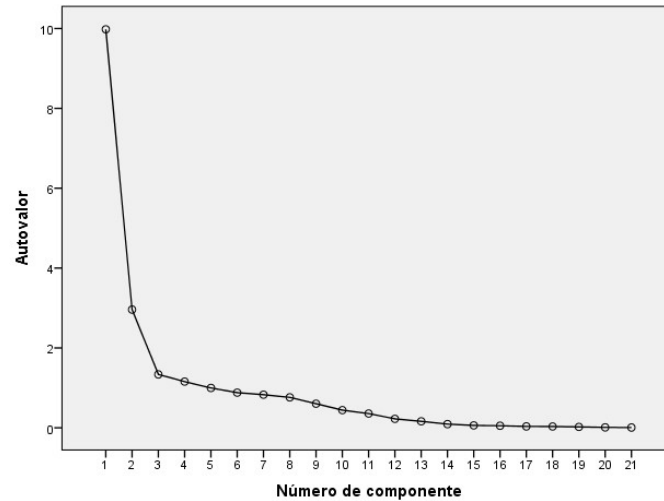
Como un soporte a la revisión de los autovalores se analizan también los puntos de la gráfica de sedimentación que tienen autovalores mayores o iguales a 1 en el eje Y. La cantidad de puntos que cumplan esa condición representan el número de posibles factores.

5.6 Matriz de Componentes Rotado

En el proceso de extracción inicial se utilizó el criterio de autovalores mayores que 1. Con este tipo de extracción el sistema arroja 4 factores como se puede apreciar en la Figura 10. El Método de rotación utilizado es Varimax, método ortogonal que minimiza el número de variables que tiene saturaciones altas en cada factor. Simplifica la interpretación de los factores optimizando la solución por columna.

Luego de revisar cuantas variables quedan asociadas por factor, se puede evidenciar que no hay factores triviales. Las variables Edad, CambioTrabajo y desocupado se excluyen debido a que tienen comunalidad baja y la carga factorial es menor a 0.5.

Debido a la eliminación de las 3 variables se hace necesario realizar nuevamente el análisis por factores. La Tabla 5 presenta los nuevos resultados obtenidos.

Figura 9. Gráfico sedimentación inicial

Fuente: Elaboración del autor

Figura 10. Valores matriz componente rotado

	Componente			
	1	2	3	4
ocupado	,950			
HorasAdicionales	,940			
OtraOcupacion	,932			
TiempoTrabajo	,919			
HorasLaborales	-,877			
Pensiones	,837			
TipoEmpleado	,824	,542		
IngresosTotal	,757			
TipoActividad	,668			
desocupado				
CambioTrabajo				
PrimaServicio		,940		
PrimaVacas		,934		
Primas		,934		
PrimaNavidad		,933		
SubTransporte		,906		
GradoEstudios			,768	
Estrato			,676	
Genero				,724
Seguridad				-,519
Edad				

Fuente: Elaboración del autor

Tabla 5. Nuevos valores análisis factorial

Indicador	Valor	Análisis
Número de variables	18	Cumple con la condición del número de variables
Alfa de Cronbach	0.818	Valor aceptable
Determinante	9,990E-14	Cercano a 0, aceptable
KMO	0.921	Valor excelente. Indica que la validación de los datos es aceptable
Chi cuadrado estimado	18983,502	El chi cuadrado estimado fue mayor que el chi tabulado lo que indica que se puede seguir con el estudio
Varianza total explicada	82.209%	La varianza acumulada es mayor al 70%, por tanto, define el número de componentes y explica el problema de manera coherente

Fuente: Elaboración del autor

Los nuevos valores de comunalidad y la nueva distribución de los factores se pueden visualizar en la Figura 11. Las comunalidades son buenas y cada factor tiene asociada dos o más variables.

Figura 11. Nuevos valores comunalidad y matriz componente rotado

	Inicial	Extracción	Componente			
			1	2	3	4
Estrato	1,000	,628				
Genero	1,000	,607				
Seguridad	1,000	,515				
GradoEstudios	1,000	,600				
TipoActividad	1,000	,630				
TiempoTrabajo	1,000	,922				
TipoEmpleado	1,000	,974				
Primas	1,000	,952				
SubTransporte	1,000	,887				
PrimaServicio	1,000	,949				
PrimaNavidad	1,000	,938				
PrimaVacas	1,000	,939				
HorasLaborales	1,000	,844				
Pensiones	1,000	,929				
OtraOcupacion	1,000	,933				
HorasAdicionales	1,000	,932				
ocupado	1,000	,979				
IngresosTotal	1,000	,639				
ocupado			,961			
HorasAdicionales			,945			
OtraOcupacion			,945			
TiempoTrabajo			,931			
HorasLaborales			-,884			
Pensiones			,850			
TipoEmpleado			,839	,518		
IngresosTotal			,771			
TipoActividad			,677			
PrimaServicio				,935		
Primas				,931		
PrimaVacas				,931		
PrimaNavidad				,930		
SubTransporte				,902		
Estrato					,784	
GradoEstudios					,743	
Genero						,753
Seguridad						-,704

Fuente: Elaboración del autor

Luego de tener el número de factores no triviales y variables asociadas a cada uno. Se procede a la normalización de los nombres de los factores de acuerdo con las relaciones entre las variables de cada factor. Al realizar el proceso de denominación para los factores se decide reubicar algunas variables debido a su naturaleza. La organización final de los factores se puede visualizar en la Tabla 6.

Tabla 6. Relación variables por factor

Factor	Variables asociadas	Porcentaje de aporte
Condiciones Laborales – F1	<ol style="list-style-type: none"> 1. Ocupado – X17 2. HorasAdicionales – X16 3. OtraOcupacion – X15 4. TiempoTrabajo – X6 5. HorasLaborales – X13 6. Pensiones – X14 7. TipoEmpleado – X7 8. TipoActividad – X5 9. Seguridad – X3 	53.9%
Bonificaciones – F2	<ol style="list-style-type: none"> 1. PrimaServicio – X10 2. Primas – X8 3. PrimaVacas – X12 4. PrimaNavidad – X11 5. SubTransporte – X9 	15.5%
Condición Social – F3	<ol style="list-style-type: none"> 1. Estrato – X1 2. GradoEstudios – X4 	6.7%
Perfil salarial – F4	<ol style="list-style-type: none"> 1. Genero – X2 2. IngresoTotal – X18 	6.2%

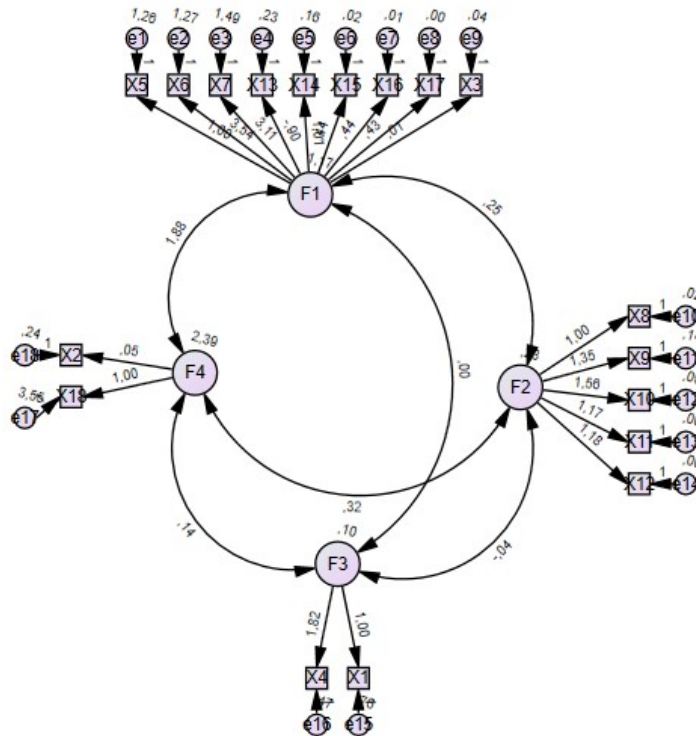
Fuente: Elaboración del autor

6. ANÁLISIS FACTORIAL CONFIRMATORIO

El propósito del Análisis factorial confirmatorio se centra en el estudio de los modelos de medida, es decir, en analizar las relaciones entre un conjunto de indicadores o variables observadas y una o más variables latentes o factores [9].

Para el caso de estudio se analiza la relación entre el conjunto de variables observables de F1: X5, X6, X7, X13, X14, X15, X16, X17 y X3. De F2: X8, X9, X10, X11, X12. F3: X1, X4. F4: X2, X18. El pat diagram que los representa se visualiza en la Figura 12.

Figura 12. Modelo de medida del estudio



Fuente: Elaboración del autor

La siguiente fase del proceso es la evaluación del ajuste del modelo. En este paso, se trata de comprobar la compatibilidad entre el modelo propuesto y la información empírica recogida, es decir, en qué medida el modelo teórico propuesto es apoyado por los datos muestrales obtenidos. Se debe examinar las estimaciones infractoras (errores estándar elevados, coeficientes estandarizados que sobrepasan la unidad, correlaciones elevadas entre dos estimaciones, entre otras). Adicionalmente, se debe tener en cuenta las medidas o índices de la calidad del ajuste modelo en su conjunto, como son las medidas de ajuste absoluto o global, las de ajuste incremental y las de parsimonia [10].

El modelo de medida cuenta con 40 variables, 18 variables observables endógenas y 22 variables no observables exógenas. En la Figura 12 se puede apreciar que hay cargas factoriales mayores a 1 en los factores 1, 2 y 3, esos valores son un indicador para analizar debido a que las cargas no deben superar ese valor. La Tabla 7 presenta los índices de calidad de ajustes del modelo con valores aceptados y los resultados obtenidos en el presente estudio.

De acuerdo con los valores de la Tabla 7 el modelo no cumple con los criterios de aceptación de los indicadores, por tal motivo, se puede decir que el modelo es inadmisibles. El alcance del análisis confirmatorio no contempla la construcción del modelo estructural, debido a que en la fase exploratoria se evidenció la ausencia de variables endógenas dependientes.

Tabla 7. Índices de bondad de ajuste del estudio

Estadístico	Criterio	Valor modelo	Análisis
Chi-cuadrado	0 a 3df	3143,013	El valor estimado no cumple con los valores aceptados.
Grados de libertad (df)	0 a 3	129	El valor estimado no cumple con los valores aceptados.
RMSEA	0 a 0.08	0.191	El valor estimado no cumple con los valores aceptados.
SRMR	0 a 0.10	0.129	El valor estimado no cumple con los valores aceptados.
NFI	0.90 a 1.00	0.836	El valor estimado no cumple con los valores aceptados.
CFI	0.95 a 1.00	0.842	El valor estimado no cumple con los valores aceptados.
GFI	0.90 a 1.00	0.691	El valor estimado no cumple con los valores aceptados.
AGFI	0.85 a 1.00	0.590	El valor estimado no cumple con los valores aceptados.

Fuente: Elaboración del autor

7. CONCLUSIONES

Respecto a la información registrada en el archivo de Excel de personas se puede decir que el proceso de depuración de los datos por parte del DANE presenta falencias, debido a que a pesar de tener en las opciones de respuesta de la mayoría de las preguntas la opción No sabe/No informa, se encontraron registros que no tenían ningún valor en la celda. Esto ocasionó que el 90% de las variables trabajadas tuvieran que ser recodificadas para corregir esos datos perdidos.

Se debe destacar la fuerte correlación entre el estrato y el grado de estudios (0.217), presenta el valor más alto con relación a todas las demás variables. Esto indica que el nivel socioeconómico influye en la formación académica de las personas en la ciudad de Montería.

Se evidenció una alta correlación entre los ingresos y tiempo de trabajo. Sin embargo, la correlación entre ingresos y grado de estudios, aunque existe de manera positiva no es la más significativa.

No se evidencia una correlación positiva entre el género y el nivel de estudios, no obstante, es representativa la correlación entre el género y el tipo de actividad.

En la fase de análisis exploratorio se identificaron 4 factores: Condiciones Laborales – F1, Bonificaciones – F2, Condición Social – F3 y Perfil salarial – F4. Es importante destacar que para el factor 4 se realizó un análisis descriptivo donde se detectó que los ingresos de los individuos del género masculino son superiores a los de su contraparte.

Al realizar el análisis factorial para este caso de estudio se puede evidenciar que las variables que influyen en mayor porcentaje en las condiciones laborales son: horas adicionales de trabajo, tener otra ocupación, el tiempo de trabajo, las horas laboradas semanalmente, el tipo de empleo, el tipo de actividad y los ingresos percibidos por su trabajo.

Del análisis de las 137 variables de la encuesta para personas del DANE se puede concluir que ninguna de las preguntas del cuestionario estaba orientada hacia recoger información relacionada con la percepción o el nivel de satisfacción de los individuos en relación con su trabajo. Por tal motivo, no fue posible trabajar con variables dependientes endógenas y así construir un modelo estructural completo.

REFERENCIAS

- [1] DANE, “Gran encuesta integrada de hogares (GEIH) Mercado laboral,” 2019. [Online]. Available: <https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo>.
- [2] G. J. Restrepo Gallego, “El Colombiano,” 2018. [Online]. Available: <https://www.elcolombiano.com/negocios/mejorar-el-trabajo-un-camello-diario-NF8632197>.
- [3] DANE, “Pobreza Monetaria y Desigualdad,” 2018.
- [4] S. De la Fuente Fernandez, “Análisis Factorial,” Madrid, 2011.
- [5] A. Fernández Aráuz, “Aplicación del análisis factorial confirmatorio a un modelo de medición del rendimiento académico en lectura,” *Rev. Ciencias Económicas*, vol. 33, no. 2, p. 39, Dec. 2015.
- [6] SurveyMonkey, “Calculadora del tamaño de muestra,” 2019. [Online]. Available: <https://es.surveymonkey.com/mp/sample-size-calculator/>.
- [7] Techlandia, “Cómo obtener una muestra aleatoria en Excel,” 2019. [Online]. Available: https://techlandia.com/obtener-muestra-aleatoria-excel-como_42965/.
- [8] M. D. Rodríguez, Á. L. G. Ariza, A. H. Pérez, and M. E. D. Mora, *Introducción al análisis estadístico multivariado aplicado. Experiencia y casos en el Caribe colombiano*. Barranquilla: Editorial Universidad del Norte, 2016.
- [9] J. Herrero, “El Análisis Factorial Confirmatorio en el estudio de la Estructura y Estabilidad de los Instrumentos de Evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14,” *Psychosoc. Interv.*, vol. 19, no. 3, pp. 289–300, 2010.
- [10] J. A. Pérez Gil, S. Chacón Moscoso, and R. Moreno Rodríguez, “Validez de constructo: El uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez,” *Psicothema*, vol. 12, no. SUPPL. 2, pp. 442–446, 2000.