

APLICACIONES ESTADÍSTICAS

*Socialización de
Experiencias*

*Memorias
2018*

*Duitama, Boyacá
UPTC-Facultad
Seccional Duitama
9 de junio*

Versión: Aplicaciones Estadísticas - 3. Socialización de Experiencias



ESPECIALIZACIÓN EN
ESTADÍSTICA



Uptc[®]
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN 3910 DE 2015 MEN / 6 AÑOS

Posgrados
Duitama



Grupo de Investigación
en Estadística

Aplicaciones Estadísticas. Socialización de Experiencias

“ISSN: 2619-2888 (En línea)”

<http://rdigitales.uptc.edu.co/memorias/>

© Universidad Pedagógica y Tecnológica de Colombia

© De cada título, su autor

© Carmen Helena Cepeda Araque, Sandra Patricia Cárdenas Ojeda, comps.

Directivas

Alfonso López Díaz

Rector

Hugo Alfonso Rojas Sarmiento

Vicerrector Académico

Enrique Vera López

Vicerrector de Investigaciones y Extensión

Adán Bautista Morantes

Decano Facultad Seccional Duitama

Hilda Lucía Jiménez Orozco

Director Escuela de Posgrados

Sandra Patricia Cárdenas Ojeda

Directora Grupo de Investigación GIE

Coordinación General

Sandra Patricia Cárdenas Ojeda

Reinaldo Alarcón Guarín

Carmen Helena Cepeda Araque

Grupo de Investigación en Estadística - GIE

Especialización en Estadística

Escuela de Posgrados

Universidad Pedagógica y Tecnológica de Colombia

Facultad Duitama

Comité Científico

Sandra Patricia Cárdenas Ojeda

Carmen Helena Cepeda Araque

Reinaldo Alarcón Guarín

Carlos Alberto Martínez Niño

Nohora Elizabeth Alfonso Bernal

Especialización en Estadística

Escuela de Posgrados

Universidad Pedagógica y Tecnológica de Colombia

Facultad Duitama

Diseño y Diagramación

Omar Velandia Castro - omarvelandia@hotmail.com

Luis Arbey Gómez Gómez – luis.gomez@uptc.edu.co

Universidad Pedagógica y Tecnológica de Colombia

Facultad Duitama

APLICACIONES
ESTADÍSTICAS

Socialización de
Experiencias

2018



ESPECIALIZACIÓN EN
ESTADÍSTICA

Contacto

Universidad Pedagógica y Tecnológica de Colombia
Facultad Seccional Duitama
Escuela de Posgrados Sede Duitama
Teléfono: (57+8) 7624431
Conmutador (57 + 8) 7605306 Ext: 2838 - 2830
Carrera 18 Calle 22 Edificio Administrativo Piso 1
Duitama - Boyacá - Colombia

www.uptc.edu.co
posgrados.duitama@uptc.edu.co

Las opiniones contenidas son responsabilidad exclusiva de sus autores y no reflejan necesariamente el pensamiento de la organización ni de la Universidad Pedagógica y Tecnológica de Colombia. Se permite la reproducción parcial o total, por cualquier medio, con la autorización expresa y escrita de los titulares del derecho de autor

APLICACIONES
ESTADÍSTICAS
*Socialización de
Experiencias*

2018



ESPECIALIZACIÓN EN
ESTADÍSTICA

PRESENTACIÓN

En calidad de coordinadora académica de la Especialización en Estadística, quisiera celebrar con ustedes este ejercicio de divulgación de los trabajos de aplicación desarrollados por los graduados de la tercera cohorte. Se ha dispuesto de este espacio para intercambiar experiencias de la aplicación de técnicas estadísticas en ámbitos como la economía, educación, agronomía, administración, ingeniería, entre otros.

La información contenida en estas memorias es el fruto de un año de intenso trabajo por parte de nuestros estudiantes, agradecemos mucho por la confianza que depositaron en nuestra Institución, y confío en que con el paso del tiempo serán recompensados por decisión de cursar la Especialización. Expresamos nuestro reconocimiento a los profesores que apoyaron los trabajos de aplicación, gracias por el profesionalismo, dedicación y buena voluntad.

Es nuestro deseo que esta publicación sea fuente de consulta para profesionales que requieren el uso de técnicas estadísticas de dependencia e interdependencia para la solución de problemas en su área de trabajo.

Carmen Helena Cepeda Araque
Coordinadora Académica
Especialización en Estadística

APLICACIONES
ESTADÍSTICAS
Socialización de
Experiencias

2018



ESPECIALIZACIÓN EN
ESTADÍSTICA

TABLA DE CONTENIDO

Variables asociables al desempeño en Inglés y comunicación escrita de las Pruebas SABER PRO TyT (Técnicas y Tecnológicas) del Cedeagro Duitama.

DIANA MILENA PINEDA ROJAS – FABIÁN DARIO BARRERA RODRÍGUEZ

diana.pinedarojas@uptc.edu.co - fdbarrera@misena.edu.co

Universidad Pedagógica y Tecnológica de Colombia-Duitama

I

Identificación morfométrica de escarabajos coprófagos (coleoptera: scarabaeidae: scarabaeinae) un ejercicio de aproximación a la taxonomía desde la estadística.

OSCAR FELIPE MORENO MANCILLA - JOSÉ DANIEL CARO CASTILLO

felipemancilla@gmail.com - dcarodaniel91@gmail.com

Universidad Pedagógica y Tecnológica de Colombia-Duitama

Fí

Formulación y validación de cuestionario para medición de las condiciones socioeconómicas de los trabajadores fabriles de las Acerías del departamento de Boyacá.

SEBASTIAN JOYA TÉLLEZ - EDGAR GABRIEL DÍAZ MELGAREJO

sebastian.joya@uptc.edu.co - edgargabriel.diaz@uptc.edu.co

Universidad Pedagógica y Tecnológica de Colombia-Duitama

G

Análisis de satisfacción y aprobación en formaciones técnicas de una Empresa Siderúrgica de Boyacá en 2017

LAURA MILENA VARGAS PÉREZ

lauramilena.vargas@uptc.edu.co

Universidad Pedagógica y Tecnológica de Colombia-Duitama

IG

Efecto de la lista de palabras de la “self-complexity task” en los puntajes de “self-complexity” en estudiantes universitarios.

LORENA SANABRIA PÉREZ

loresan288@hotmail.com

Universidad Pedagógica y Tecnológica de Colombia-Duitama

ii

Análisis del ausentismo laboral por incapacidades médicas en una empresa siderúrgica de Colombia en el año 2017.

YULI STEFANY SANDOVAL VELANDIA

yulisandoval16@hotmail.com

Universidad Pedagógica y Tecnológica de Colombia-Duitama

iii

APLICACIONES
ESTADÍSTICAS

Socialización de
Experiencias

2018



ESPECIALIZACIÓN EN
ESTADÍSTICA



VARIABLES ASOCIABLES AL DESEMPEÑO EN INGLÉS Y COMUNICACIÓN ESCRITA DE LAS PRUEBAS SABER PRO TyT (TÉCNICAS Y TECNÓLOGICAS) DEL CEDEAGRO DUITAMA

Especialización en Estadística

DIANA MILENA PINEDA ROJAS^{1,a}, FABIÁN DARÍO BARRERA RODRÍGUEZ^{2,b}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

En este artículo se encuentra la descripción y análisis de las pruebas SABER TyT de Tecnología en Procesamiento de Alimentos del Servicio Nacional de Aprendizaje SENA, Centro de Desarrollo Agropecuario y Agroindustrial CEDEAGRO- Duitama, I y II periodo 2016 y I periodo 2017 para el desempeño en los módulos de inglés y comunicación escrita, como técnica estadística se empleó un modelo de Regresión Logística Ordinal (RLO).

Palabras clave: SABER TyT, Regresión Logística Ordinal, Niveles de desempeño.

Abstract

In this article we find the description and analysis of the SABER TyT tests of Food Processing Technology of the National Service of Learning SENA, Agroindustrial Development Center and Agroindustrial CEDEAGRO- Duitama, I and II period 2016 and I period 2017 for the performance in the modules of English and written communication, as a statistical technique an Ordinal Logistic Regression model (RLO) was used

Key words: SABER TyT, Ordinal Logistic Regression, Performance levels.

1. Introducción

El Ministerio de Educación Nacional de Colombia con el ánimo de mejorar la calidad académica de los programas tecnológicos, y con el fin de ejercer su inspección y vigilancia instauró la presentación obligatoria de un examen estandarizado a nivel nacional conocido actualmente como SABER PRO TyT. Dicho examen pretende evaluar las competencias necesarias para el adecuado desempeño profesional.

De acuerdo con Duarte, Peña y Bustos (2015) en la actualidad el Instituto Colombiano de Fomento de la Educación Superior -ICFES- es la entidad responsable de la evaluación de la calidad de la educación básica por medio de las pruebas Saber tercero, quinto y noveno, al igual que la evaluación de la educación media (Saber 11) y la educación superior (Saber Pro).

^aEstudiante de Especialización en Estadística. E-mail: diana.pinedarojas@uptc.edu.co

^bEstudiante de Especialización en Estadística. E-mail: fdbarrera@misena.edu.co

Por tanto, en este estudio se pretendió identificar las variables asociables al desempeño en inglés y comunicación escrita de las pruebas saber TyT en el programa de Tecnología en Procesamiento de Alimentos del Servicio Nacional de Aprendizaje SENA, Centro de Desarrollo Agropecuario y Agroindustrial CEDEAGRO-Duitama. En particular el estudio se centra en las pruebas del año 2016 y primer periodo de 2017, se tuvieron en cuenta 94 observaciones de la base de datos del FTP (File Transfer Protocol) del ICFES.

El Servicio Nacional de Aprendizaje SENA busca contribuir con el mejoramiento de la educación en Colombia en los niveles tecnológicos, y teniendo en cuenta esta necesidad surgió esta investigación, puesto que identificando las fortalezas y falencias que presentan las competencias evaluadas se puede contribuir en planes de mejora que incidan en la calidad de formación de los aprendices.

En ese sentido, es preciso resaltar que la dirección general de la entidad expidió la Resolución 1016 de 2013 por la cual establece que los aprendices del SENA del nivel Tecnólogo deberán presentar el Examen de Calidad de la Educación Superior, Saber Pro, como requisito adicional para obtener el título y aplica para todos los aprendices que están en séptimo y octavo trimestre de su proceso de formación o que hayan alcanzado el 75 % del total de su formación presencial o virtual.

Así mismo, los lineamientos del plan de acción 2018 del SENA, recomiendan “que las Regionales y Centros de formación establecerán con base en los resultados reportados por el ICFES las fortalezas y falencias que presentan las competencias evaluadas y así ejecutar planes de mejora a los diseños y programas que redunden en la calidad de la formación”

Con relación a algunos antecedentes investigativos se relacionan los estudios de Duarte, Peña y Bustos (2015) quienes a través de un modelo de elección binaria tipo logit contrastaron la incidencia de variables explicativas sobre los tres módulos genéricos dispuestos en las pruebas: lectura crítica, razonamiento cuantitativo e inglés. La conclusión del este estudio mostró que las variables con mayor incidencia positiva en los más altos niveles de desempeño son: el género, el estado civil, el lugar de residencia, la vinculación laboral, tener o no personas a cargo, el estrato socio-económico, el origen de la institución en donde cursó el bachillerato (oficial o privada), y la tenencia de conexión a internet.

Por otra parte los estudios de Rico, Hernández y Alonso (2014) relacionan el empleo del modelo de regresión logística ordinal (RLO) para la predicción del rendimiento académico, con el objeto de realizar intervención oportuna en los estudiantes y de esta manera reducir el índice de fracaso escolar en una asignatura de segundo año de Ingeniería Industrial en la Habana, Cuba. Esta misma técnica estadística ha sido empleada por Gonzalez y Cepeda (2016) para evaluar del desempeño académico en matemáticas en el colegio Mariano Ospina Pérez de la ciudad de Bogotá. El modelo de RLO fue utilizado en este estudio para medir el desempeño de los aprendices del CEDEAGRO Duitama del en los módulos genéricos de las pruebas saber TyT anteriormente descritos.

2. Referente Conceptual

En esta sección se presentan algunos conceptos relacionados con la aplicación de las pruebas SABER TyT, las características de los módulos de los componentes genéricos: desempeño en Inglés y Comunicación escrita; y finalmente se abordará lo relacionado con el modelo de Regresión Logística Ordinal (RLO).

2.1. Pruebas saber PRO

En el SENA se han desarrollado algunos estudios previos como los de Duarte, Peña y Bustos (2015) quienes analizaron el desempeño de los aprendices SENA en el orden nacional en las pruebas saber TyT 2012 con relación a algunas variables sociodemográficas institucionales asociadas a los estudiantes; De manera general los resultados mostraron una marcada incidencia entre la mayoría de las variables sociodemográficas e institucionales empleadas en el estudio sobre los puntajes alcanzados por los aprendices SENA en las competencias de lectura crítica, razonamiento cuantitativo e inglés en las pruebas Saber Pro 2012. Con relación a la edad se concluyó que las edades tempranas hacen que el aprendiz dedique menos tiempo para el aprendizaje; en cuanto a los Centros de formación en los que la mayor proporción son aprendices del género femenino se requiere implementar estrategias para el refuerzo en módulo de inglés y de razonamiento

cuantitativo; respecto al estado civil los resultados muestran que los aprendices solteros tienen más altos niveles de desempeño.

De otra manera la distribución de los recursos económicos entre los departamentos incide en la obtención de medios y condiciones más adecuados para brindar una mejor calidad de educación, generando una visible desigualdad de posibilidades entre los aprendices, reflejada esta, en el nivel de competencia alcanzado al término de la formación profesional.

Para el desempeño de los alumnos en la Educación Superior se relacionan los estudios de Sánchez (2017), quien desarrolló un trabajo investigativo asociado con el análisis descriptivo de los resultados de las pruebas saber para estudiantes de la UPTC del periodo 2012 a 2015.

De acuerdo con el Decreto 3963 de octubre de 2009, son objetivos del SABER PRO los siguientes:

- Comprobar el grado de desarrollo de las competencias de los estudiantes próximos a culminar los programas académicos de pregrado que ofrecen las instituciones de educación superior.
- Producir indicadores de valor agregado de la educación superior en relación con nivel de competencias de quienes ingresan a este nivel; proporcionar información para la comparación entre programas, instituciones y metodologías, y mostrar su evolución en el tiempo.
- Servir de fuente de información para la construcción de indicadores de evaluación de la calidad de los programas e instituciones de educación superior y del servicio público educativo, que fomenten la cualificación de los procesos institucionales y la formulación de políticas, y soporten el proceso de toma de decisiones en todos los órdenes y componentes del sistema educativo.

Del componente genérico de la prueba Saber Pro se tuvo en cuenta:

Desempeño en inglés: Este módulo evalúa la competencia para comunicarse efectivamente en inglés mediante preguntas de selección múltiple con única respuesta. Estas competencias, alineadas con el Marco Común Europeo, permiten clasificar a los examinados en cuatro niveles de desempeño A1, A2, B1, B2.

Nivel A1: • El estudiante es capaz de comprender y utilizar expresiones cotidianas de uso muy frecuente, así como frases sencillas destinadas a satisfacer necesidades de tipo inmediato.

- Puede presentarse él mismo y ante otros, pedir y dar información personal básica sobre su domicilio, sus pertenencias y las personas que conoce.
- Puede relacionarse de forma elemental siempre que su interlocutor hable despacio y con claridad y esté dispuesto a cooperar.

Nivel A2: • El estudiante es capaz de comprender frases y expresiones de uso frecuente relacionadas con áreas de experiencia especialmente relevantes (información básica sobre él mismo y su familia, compras, lugares de interés, ocupaciones, etc.).

- Sabe comunicarse a la hora de llevar a cabo tareas simples y cotidianas que no requieran más que intercambios sencillos y directos de información sobre cuestiones conocidas o habituales.
- Sabe describir en términos sencillos aspectos de su pasado y su entorno, así como cuestiones relacionadas con sus necesidades inmediatas.

Nivel B1: • El estudiante es capaz de comprender los puntos principales de textos claros y en lengua estándar, si tratan cuestiones conocidas, ya sea en situaciones de trabajo, de estudio o de ocio.

- Sabe desenvolverse en la mayoría de las situaciones que pueden surgir durante un viaje por zonas donde se utiliza la lengua.
- Es capaz de producir textos sencillos y coherentes sobre temas que le son familiares o en los que tiene un interés personal.
- Puede describir experiencias, acontecimientos, deseos y aspiraciones, así como justificar brevemente sus opiniones o explicar sus planes.

- Nivel B2:**
- El estudiante es capaz de entender las ideas principales de textos complejos que traten de temas concretos abstractos, incluso si son de carácter técnico, siempre que estén dentro de su especialización.
 - Puede relacionarse con hablantes nativos con un grado suficiente de fluidez y naturalidad, de modo que la comunicación se realice sin esfuerzo por ninguno de los interlocutores.
 - Puede producir textos claros y detallados en torno a temas diversos, así como defender un punto de vista sobre temas generales indicando los pros y los contras de las distintas opciones.

Desempeño en comunicación escrita: Los niveles de desempeño consisten en una descripción cualitativa de las habilidades y conocimientos de los estudiantes que podrían tener si se ubican en determinado nivel. El resultado en el módulo de escritura se interpreta de acuerdo con la descripción de los niveles de desempeño que se presenta a continuación:

NIVEL 1: Los estudiantes que se ubican en este nivel pueden escribir textos que:

- Responden a la pregunta planteada en la tarea.
- Expresan ideas desarticuladas entre sí, que no dan cuenta de un planteamiento.
- Presentan dificultades en el manejo de la convención (sintaxis, escritura de las palabras, segmentación, omisión de letras, etc.) que no permiten la comprensión de sus ideas.

NIVEL 2: Además de lo descrito en el nivel anterior, los estudiantes que se ubican en este nivel pueden escribir texto que:

- Evidencian un planteamiento o posición personal para cumplir una intención comunicativa.
- Presentan algunas fallas en su estructura y organización, que hacen que estos carezcan de unidad semántica.
- Muestran algunas contradicciones, digresiones o repeticiones que afectan la coherencia del texto.
- Presentan algunos errores en el manejo de la convención, aunque estos no afectan la comunicación de sus ideas.

NIVEL 3: Además de lo descrito en los niveles anteriores, los estudiantes que se ubican en este nivel pueden escribir texto que:

- Emplean una estructura básica con un inicio, un desarrollo y un cierre.
- Se desarrollan en un mismo eje temático, de modo que este alcanza unidad.
- Presentan argumentos suficientemente desarrollados para apoyar la posición planteada.
- Tienen unidad semántica, aunque pueden incluir información innecesaria que afecta la fluidez.
- Hacen un buen uso del lenguaje, aunque pueden identificarse errores de puntuación y fallas de cohesión local.

NIVEL 4: Además de lo descrito en los niveles anteriores, los estudiantes que se ubican en este nivel pueden escribir texto que:

- Muestran diferentes perspectivas sobre el tema que complejizan el planteamiento y permiten cumplir satisfactoriamente con el propósito comunicativo propuesto en la pregunta.
- Presentan recursos semánticos, pragmáticos y estilísticos que apoyan el planteamiento del texto.
- Hacen uso adecuado de signos de puntuación, referencias gramaticales, conectores, entre otros mecanismos cohesivos, que garantizan la coherencia y fluidez del texto.

2.2. Regresión Logística Ordinal (RLO)

La regresión logística es una técnica estadística multivariante que permite estimar la relación existente entre una variable dependiente no métrica, en particular dicotómica y un conjunto de variables independientes métricas o no métricas. De acuerdo con Agresti (2013) citado en Balaguer, Sinha y Goitía (2009) La regresión logística en su forma más simple, es decir, con una respuesta binaria, propone que el logaritmo de la “razón de probabilidad” (odds según su denominación en inglés), entendida como el cociente entre la probabilidad de éxito y la de fracaso en un ensayo de Bernoulli, es igual a una función lineal en los parámetros, denominada usualmente predictora lineal. Hosmer y Lemeshow (2000) establecen que se pueden emplear tres modelos de regresión logística; Modelo de categoría adyacente, modelo de razón continua y modelo de la razón de odds, siendo este último el utilizado en este caso, debido a que las variables de respuesta desempeño en inglés y comunicación escrita son ordinales y presentan distintas categorías.

En el caso particular de la RLO se utiliza una función de unión para relacionar de forma lineal a las variables explicativas con la razón de probabilidad entre la probabilidad acumulada hasta la categoría i de la variable ordinal, y la probabilidad que la variable tome un valor mayor que la categoría i (Agresti 2002) ;Hosmer y Lemeshow (2000).

En la RLO las funciones de unión más empleadas son la Logit y la Cloglog Agresti (2002). McCullagh (1980) plantea que la unión con Logit es más adecuada para analizar datos ordinales, cuya distribución de frecuencia es uniforme a lo largo de todas las categorías, mientras que la unión Cloglog es preferible para analizar datos categóricos, cuyas categorías de mayor valor son las más probables. Para el caso puntual de este estudio, los valores de la variables ordinal estan representados por los desempeños A1, A2, B1, B2 en inglés y nivel 1, nivel 2, nivel 3, nivel 4 en comunicación escrita, que pueden obtener los aprendices SENA en las pruebas Saber TyT, así las cosas es conveniente considerar que las categorías más altas presentan un mayor grado de complejidad dentro del módulo de competencias genéricas evaluadas por el ICFES, es decir, estos desempeños son los menos probables.

De acuerdo con Agresti (2013) la expresión de la función Logit para la RLO es la siguiente:

$$\begin{aligned} \text{logit}[P(Y \leq j|x)] &= \log \frac{P(Y \leq j|x)}{1 - P(Y \leq j|x)} \\ &= \log \frac{\pi_1(x) + \dots + \pi_j(x)}{\pi_{j+1}(x) + \dots + \pi_J(x)}, \quad j = 1, \dots, J - 1 \end{aligned} \quad (1)$$

Cada logit acumulativo usa todas las categorías de respuesta J

Un modelo para $\text{logit}[P(Y \leq j)]$ solo es un modelo logístico ordinario para una respuesta binaria en la que las categorías 1 hasta j forman un resultado y las categorías $j+1$ hasta J forman el segundo. Un modelo que usa simultáneamente todas las categorías ($J - 1$) en logit acumulativo es un único modelo más simple (parsimonioso)

$$\text{logit}[P(Y \leq j|x)] = \alpha_j + \beta^T x, \quad j = 1, \dots, J - 1 \quad (2)$$

Cada logit acumulativo tiene su propio intercepto. Los $\{\alpha_j\}$ aumentan en j , porque $[P(Y \leq j|x)]$ aumenta en j para x fijo y el logit es una función creciente de $[P(Y \leq j|x)]$

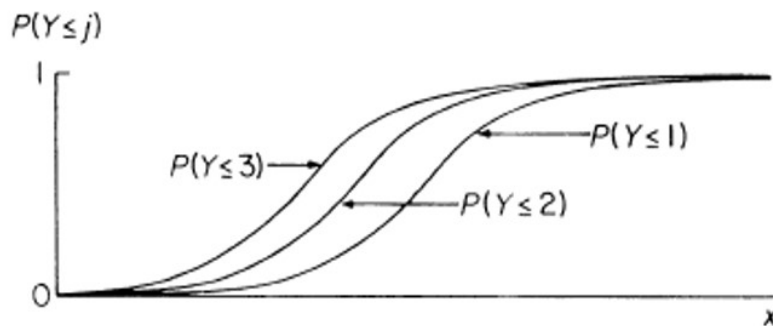


FIGURA 1: Representación de las probabilidades acumuladas en el modelo de razón de odds
Agresti (2007)

Este modelo asume el mismo efecto β para cada logit. Para un solo predictor x , la figura representa el modelo cuando $J = 4$ para j fijo, la curva de respuesta es una curva de regresión logística para una respuesta binaria con resultado ($Y \leq j$) y ($Y > j$). Las curvas para $j = 1, 2, y 3$ tienen la misma forma. Ellas comparten exactamente la misma tasa de incremento o disminución, pero horizontalmente están distanciadas una de la otra.

El modelo logit satisface:

$$\begin{aligned} \text{logit}[P(Y \leq j|x_1)] - \text{logit}[P(Y \leq j|x_2)] &= \log \frac{P(Y \leq j|x_1)}{1 - P(Y \leq j|x_1)} - \log \frac{P(Y \leq j|x_2)}{1 - P(Y \leq j|x_2)} \\ &= \log \frac{P(Y \leq j|x_1)/[1 - P(Y \leq j|x_1)]}{P(Y \leq j|x_2)/[1 - P(Y \leq j|x_2)]} = \beta^T(x_1 - x_2) \quad (3) \end{aligned}$$

Una probabilidad acumulada de la razón de odds es llamada razón de odds acumulada. Los odds para obtener una respuesta $\leq j$ en $x = x_1$ son $\exp[\beta^T(x_1 - x_2)]$ veces la probabilidad en $x = x_2$. El log de la razón de odds acumulada es proporcional a la distancia entre x_1 y x_2 . La misma constante de proporcionalidad aplica para cada logit. debido a esta propiedad, a menudo el modelo (2) es llamado modelo de odds proporcional McCullagh (1980) citado por Agresti (2013). Con un solo predictor la razón de odds acumulada es igual a e^β cuando $x_1 - x_2 = 1$

Para estimar los coeficientes del modelo y los errores estándar se utiliza la estimación por la función de verosimilitud (likelihood). El estimador máximo verosimil (ML) se puede obtener por el método iterativo de Newton-Raphson Agresti (2013).

$$\begin{aligned} \prod_{i=1}^n \left[\prod_{j=1}^J \pi_j(x_i)^{y_{ij}} \right] &= \prod_{i=1}^n \left\{ \prod_{j=1}^J [P(Y \leq j|x_i) - P(Y \leq j-1|x_i)]^{y_{ij}} \right\} \\ &= \prod_{i=1}^n \left\{ \prod_{j=1}^J \left[\frac{\exp(\alpha_j + \beta^T x_i)}{1 + \exp(\alpha_j + \beta^T x_i)} - \frac{\exp(\alpha_{j-1} + \beta^T x_i)}{1 + \exp(\alpha_{j-1} + \beta^T x_i)} \right]^{y_{ij}} \right\} \quad (4) \end{aligned}$$

Lo que significa evaluar si en conjunto las variables regresoras tienen un efecto estadísticamente significativo sobre la variable dependiente. Con este fin se emplea el estadístico G, calculado según la fórmula:

$$G = \frac{-2 \ln(\text{probabilidad de la muestra sin variables})}{\text{probabilidad de la muestra con variables}} \quad (5)$$

Este estadístico sigue distribución Ji-cuadrado, con tantos grados de libertad como variables independientes existan McCullagh (1980) citado por Agresti (2002). Cuando las predicciones de la variable dependiente que se hacen con el modelo que incluye todas las variables independientes superan las predicciones que se

realizan sin considerar éstas, el valor de G tiende a ser grande y conlleva a concluir que al menos una de las variables regresoras tiene efecto sobre la variable dependiente, y, por lo tanto, que la probabilidad de ocurrencia de los valores que representan esta variable varía para alguna de las combinaciones de valores de las variables independientes Agresti (2002). La significancia individual de cada variable independiente, generalmente, se analiza a través de la prueba de Wald, la cual se basa en la significancia del coeficiente estimado para cada variable Agresti (2002). El estadístico utilizado es:

$$Z_{Wald} = \frac{b_j}{ES(b_j)} \quad (6)$$

donde b_j es el coeficiente de regresión estimado para la variable independiente j .

Bajo la hipótesis de que el coeficiente poblacional $\beta_j = 0$ para la variable j , la razón entre la estimación de este coeficiente (b_j) y el error estándar de esta estimación [$ES(b_j)$], sigue una distribución normal estándar. Mayores valores de este estadístico indican que el coeficiente b_j es distinto de cero, y, por ende, que la variable independiente tiene efecto sobre la probabilidad de ocurrencia de los valores de la variable dependiente.

Además de mirar si las variables independientes tienen efecto sobre la variable dependiente, es necesario conocer si el modelo que se obtiene tiene buen ajuste. siguiendo a Rico, Hernández y Alonso (2014) Usualmente para evaluar el ajuste se construye una tabla de contingencia cuyas filas representan los valores de la variable objeto de estudio y cuyas columnas representan las posibles combinaciones de valores de las variables independientes. Cuando alguna variable regresora es continua la tabla se construye considerando los ajustes recomendados por Hosmer y Lemeshow (2000). El empleo de una prueba de bondad de ajuste para comparar en cada celda de la tabla los valores observados y los valores predichos según el modelo es útil para valorar la calidad de ajuste. Si la frecuencia predicha para las combinaciones según el modelo, difiere significativamente de la frecuencia con la cual ocurren realmente los valores en estas combinaciones, existe evidencia de falta de ajuste. Para comparar estas frecuencias generalmente se calcula el estadístico chi cuadrado de Pearson según la fórmula:

$$x^2 = \sum_{i=1}^k \sum_{l=1}^m \frac{(y_{il} - m_l p_{il})^2}{m_l p_{il} (1 - p_{il})} \quad (7)$$

Tasa de clasificaciones correctas: Para cuantificar la bondad del ajuste global del modelo se dispone también de otra medida como es la tasa de clasificaciones correctas. Es decir, a partir del modelo ajustado, se clasifica cada observación en la categoría más probable, construyendo así una matriz de clasificación observados-predichos y se utiliza el porcentaje de clasificaciones correctas como una medida de la calidad de predicción, del mismo modo que se hace en el análisis discriminante. Se define como la proporción de individuos clasificados correctamente por el modelo y se calcula como el cociente entre el número de observaciones clasificadas correctamente y el tamaño muestral N. Un individuo es clasificado correctamente por el modelo cuando su valor observado de la variable respuesta. $Y(Y_1, Y_2, \dots, Y_k)$ coincide con su valor estimado por el modelo DUEÑAS (2006)

Calidad de ajuste: De acuerdo con DUEÑAS (2006) En los modelos de regresión logística binaria, la calidad del ajuste se mide mediante coeficientes de determinación conocidos como Pseudo- R^2 , para la regresión logística multinomial también se utilizan estos coeficientes. Los más usados son: Mc-Fadden, el de Cox-Snell y el de Nagelkerke. a continuación se ve cómo se calculan cada uno de ellos.

Coficiente pseudo- R^2 de Mc-Fadden. Si tenemos $\Lambda = -2 \ln(V)$, identificamos por Λ_0 el valor inicial de esta función, es decir el mínimo Λ bajo el modelo nulo dado sólo por un término constante y por Λ_f el mínimo de Λ bajo el modelo ajustado con todos los parámetros, obtenemos la siguiente expresión del pseudo- R^2 de Mc-Fadden:

$$R_{MF}^2 = 1 - \frac{\Lambda_f}{\Lambda_0} \quad (8)$$

Siendo su rango teórico de valores $0 \leq R_{MF}^2 \leq 1$, pero muy raramente su valor se aproxima a 1. Suele considerarse una buena calidad de ajuste cuando $0.2 \leq R_{MF}^2 \leq 0.4$ y excelente para valores superiores.

Coficiente pseudo- R^2 de Cox-Snell. En este caso se utiliza directamente la función de verosimilitud V , y no la función de Λ . Por lo que si denotamos por $V_0 = \exp\left(\frac{-\Lambda_0}{2}\right)$ el máximo de verosimilitud bajo el

modelo nulo dado sólo por un término constante y por $V_f = \exp\left(\frac{-\Lambda_f}{2}\right)$ el máximo de verosimilitud bajo el modelo ajustado con todos los parámetros, definimos el coeficiente pseudo- R^2 de Cox-Snell como:

$$R_{CS}^2 = 1 - \left(\frac{V_0}{V_f}\right)^{\frac{2}{N}} = 1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right) \quad (9)$$

El rango teórico de valores para el coeficiente es $0 \leq R_{CS}^2 \leq 1 - (V_0)^{\frac{2}{N}}$, lo que le hace poco interpretable al depender de V_0 . Ya que puede ser próximo a cero cuando hay pocos datos. Por ello es preferible utilizar el siguiente coeficiente como medida de bondad de ajuste.

Coficiente pseudo- R^2 de Nagelkerke. Viene dado por la siguiente expresión:

$$R_N^2 = \frac{R_{CS}^2}{1 - (V_0)^{\frac{2}{N}}} = \frac{1 - \exp\left(\frac{\Lambda_f - \Lambda_0}{N}\right)}{1 - \exp\left(\frac{-\Lambda_0}{N}\right)} \quad (10)$$

Y en este caso, su rango de valores es $0 \leq R_N^2 \leq 1$, por lo que puede interpretarse del mismo modo que el coeficiente de determinación de la regresión lineal clásica, aunque es más difícil que alcance valores cercanos a 1.

Por último decir que, para comparar modelos de regresión logística multinomial con diferente número de variables predictoras suelen introducirse coeficientes pseudo- R^2 ajustados. El más conocido es el de McFadden, definido como:

$$Adj - R_{MF}^2 = 1 - \frac{0.5\Lambda_f + n + 1}{0.5\Lambda_0 + 1} \quad (11)$$

Siendo n el número de variables predictoras.

En la expresión se forman m combinaciones con los valores de las variables explicativas y se tiene en cuenta que la variable respuesta tiene k niveles o categorías de manera que: y_{il} es la frecuencia observada de la i -ésima categoría de la variable dependiente en la l -ésima combinación de valores de las variables explicativas. p_{il} es la probabilidad estimada con el modelo para la i -ésima categoría de la variable dependiente en la l -ésima combinación de valores de las variables independientes. m_l es la cantidad de elementos en la l -ésima combinación de valores de las variables explicativas. Mientras mayor es el valor del estadístico x^2 mayor sospecha de falta de ajuste. Si finalmente se concluye la existencia de relación entre las variables explicativas y la dependiente, y si la ecuación lograda presenta buen ajuste, entonces se pueden hacer otros análisis, por ejemplo, para obtener la razón de probabilidad acumulada de la categoría i de la variable dependiente para determinados valores de las independientes, se despeja esta razón de la función logarítmica de forma que:

$$\frac{P(\text{Valor sea} \leq \text{categoría } i / \text{Valores de } x)}{P(\text{Valor sea} > \text{categoría } i / \text{Valores de } x)} = e^{\alpha_i + \beta x} \quad (12)$$

Con el término “valor” en la expresión (12) se hace referencia a cualquier valor que pueda tomar la variable dependiente. De la misma expresión se deriva que:

$$P(\text{Valor sea} \leq \text{categoría } i / \text{Valores de } x) = \frac{e^{\alpha_i + \beta x}}{1 + e^{\alpha_i + \beta x}} \quad (13)$$

y de (13) se deduce que:

$$P(\text{Valor sea} = \text{categoría } i / \text{Valores de } x) = P(\text{Valor sea} \leq \text{categoría } i) - P(\text{Valor sea} \leq \text{categoría } i - 1) \quad (14)$$

La anterior expresión es de gran importancia puesto que posibilita estimar a través de ella y dado un conjunto de valores de las variables regresoras, la probabilidad que la dependiente tome cada uno de sus valores. También suele calcularse el odds ratio (ratio de la razón de probabilidad) que genera el cambio en cada una de las variables independientes. El odds ratio de la variable independiente x evalúa la relación entre la razón

de probabilidad asociada a la categoría i cuando $x = x_2$, y la razón de probabilidad asociada a la categoría i cuando $x = x_1$ Rico, Hernández y Alonso (2014).

$$\text{odds ratio} = \frac{P(Y \leq i|X = X_2)/P(Y > i|X = X_2)}{P(Y \leq i|X = X_1)/P(Y > i|X = X_1)} \quad (15)$$

Las categorías de la variable respuesta se ven afectadas de igual forma por una determinada variable explicativa, de esta manera el odds ratio es usado para identificar el efecto de las variables explicativas sobre la variable respuesta. Según los planteamientos de Agresti (2002), Si éste es igual a uno indica que la variable explicativa no tiene efecto. Si es menor que uno, lo cual sucede cuando el coeficiente de la variable regresora es negativo, indica que, si las otras variables explicativas permanecen constante, los cambios en la variable explicativa analizada incrementan la probabilidad de obtener categorías de mayor valor en la variable objeto de estudio Hosmer y Lemeshow (2000). Valores de odds ratio mayores que uno muestran que las variaciones en la variable independiente disminuyen la probabilidad de obtener categorías de mayor valor de la dependiente Gonzalez y Cepeda (2016).

3. Diseño Metodológico

En esta investigación se utilizó el enfoque cuantitativo y se empleó como fuente primaria los datos del ICFES, para los resultados de las pruebas Saber TyT de los aprendices de Tecnología en Procesamiento de Alimentos del Servicio Nacional de Aprendizaje SENA, Centro de Desarrollo Agropecuario y Agroindustrial-CEDEAGRO Duitama, con una muestra de 126 aprendices participantes, de los cuales se obtuvieron como observaciones validas 94.

Los datos fueron tomados de la página del ICFES ICFES (2017) asignando un usuario y contraseña que permitió ingresar al FTP (File Transfer Protocol). Se descargó el documento diccionario de variables saber TyT para su interpretación y codificación; una vez obtenida la información se realizó un análisis para identificar las variables a trabajar que tuvieran coincidencia en los tres reportes del ICFES, logrando así la depuración y unificación para un total de 94 observaciones en el periodo de tiempo I II periodos de 2016 y I periodo de 2017. A continuación, se realizó un estudio univariado con el objeto de identificar que variables pudieran ser significativas para aplicar en el modelo de regresión logística ordinal.

Una vez se tiene la muestra de 94 estudiantes, se procedió a identificar las variables y el tipo de variable. Para el caso de las variables cuantitativas, tales como, edad, se calcularon algunas estadísticas de tendencia central, posición, forma y apuntamiento.

Para todas las variables se hace el análisis descriptivo. El procesamiento de los datos se llevó a cabo en el software libre R(R Core Team 2015). Finalmente la organización del documento.

4. Desarrollo

A continuación se presenta un análisis descriptivo de las variables estudiadas, se aprecia la distribución univariada de las variables independientes como la dependiente, las unidades en las que se miden o los valores codificados que toman, a fin de tener en cuenta cada una de las características de la variable, así como su naturaleza.

La aplicación de la regresión logística ordinal para obtener un modelo que permitió pronosticar los resultados probabilísticos de los desempeños de inglés y comunicación escrita, la cual se imparte en el I,II periodo 2016 y I periodo 2017 del CEDEAGRO Duitama. La variable dependiente es la calificación del aprendiz en cada desempeño, cuya escala es ordinal, los valores que toman son: “A1”, “A2”, “B1”, “B2” para inglés e “Insuficiente”, “Mínimo”, “Satisfactorio”, “Avanzado” para comunicación escrita.

Variable	Unidades/ Valores que toma/ Codificación	Descriptivo
Desempeño en inglés	A1 A2 B1 B2	A1: 53 aprendices (57 %) A2: 35 aprendices (37 %) B1: 6 aprendices (6 %) B2: 0 aprendices (0 %)
Desempeño en Comunicación escrita	Nivel 1: Insuficiente Nivel 2: Mínimo Nivel 3: Satisfactorio Nivel 4: Avanzado	Nivel 1: 11 aprendices (12 %) Nivel 2: 50 aprendices (53 %) Nivel 3: 21 aprendices (22 %) Nivel 4: 12 aprendices (13 %)
Edad	20 - 28 años 29 - 37 años 38 - 46 años 47 - 55 años	20 - 28: 84 aprendices (90 %) 29 - 37: 5 aprendices (5 %) 38 - 46: 3 aprendices (3 %) 47 - 55: 2 aprendices (2 %)
Género	Masculino: M Femenino: F	M=Masculino= 22 aprendices (23 %) F=Femenino= 72 aprendices (77 %)
Educación de la madre	Primaria: P Secundaria: S	P=Primaria= 45 (48 %) S=Secundaria= 38 (40 %) NA= 11 (12 %)
Educación del padre	Primaria: P Secundaria: S	P=Primaria=54 (57 %) S=Secundaria=29 (31 %) NA=11 (12 %)
Internet	Cuenta con Internet: SÍ No cuenta con Internet: NO	SÍ: 17 (18 %) NO: 66 (70 %) NA: 11 (12 %)

TABLA 1: Descripción de las variables estudiadas

De las pruebas SABER TyT presentadas por aprendices de la Tecnología en Procesamiento de Alimentos del Servicio Nacional de Aprendizaje SENA, Centro de Desarrollo Agropecuario y Agroindustrial CEDEAGRO- Duitama, I y II periodo 2016 y I periodo 2017 se puede afirmar que:

Con relación al desempeño en Inglés 53 aprendices que corresponde al (57%) se situaron en la categoría A1, 35 aprendices en A2 con un 37%, 6 aprendices equivalente al 6% en nivel B1 y ningún aprendiz en el nivel B2.

Para el módulo desempeño en comunicación escrita el 12% correspondiente a 11 aprendices se ubicaron en nivel 1 insuficiente, el 53% equivalente a 50 aprendices en el nivel 2 mínimo, el 22%, es decir, 21 aprendices en el nivel 3 satisfactorio y finalmente 12 aprendices que equivale a un 13% en el nivel 4 avanzado.

El 77% correspondiente a 72 aprendices que presentaron la prueba fueron mujeres, mientras que el 23% equivalente a 22 aprendices fueron hombres.

Respecto de la variable educación de la madre de los aprendices el 48% tienen nivel educativo primaria, el 40% cuentan con educación secundaria y un 12% no dieron respuesta al ítem indagado por parte del ICFES

Para la educación del padre el 57% cuentan con educación primaria, el 31% con educación secundaria, el 12% de los aprendices no dio respuesta a este ítem consultado por el ICFES, finalmente el 70% de los aprendices no cuentan con el servicio de internet que equivale a 66 personas y un 18% correspondiente a 17 aprendices sí disponen de este servicio, 11 personas no dieron respuesta a esta pregunta.

Mínimo	Máximo	Media	Q1	Q2	Q3	SD	CV	Asimetría	Kurtosis
20	54	23.44	20	21	23	6.1427	26.1987	2.9772	12.4944

TABLA 2: Algunas estadísticas de la variable edad

El 50% de los aprendices tiene una edad igual o inferior a 21 años de edad. Es notorio que los aprendices de edades entre 20 y 28 años conforman la mayor parte de aprendices que presentaron las pruebas Saber Pro, ya que corresponde al 90% del total de aprendices que la presentarán.

Modelo del desempeño en inglés

Variable	Estimación	Error estándar	Razón odds	Intervalo de confianza	
Edad	-0.02436	0.03852	0.9759	0.9050	1.0525
Género [T.M]	0.43419	0.53233	1.5437	0.5438	4.3822
Educación de la madre [T.Secundaria]	0.30192	0.48482	1.3525	0.5229	3.4979
Educación del padre [T.Secundaria]	-0.12137	0.51089	0.8857	0.3254	2.4108
Internet [T.SI]	0.12385	0.54983	1.1318	0.3853	3.3251

TABLA 3: Interpretación del modelo de desempeño en inglés

	Value	Std. Error	t value
A1 A2	-0.0774	0.9569	-0.0809
A2 A3	2.4335	1.0310	2.3603

Resid. Deviance	141.0152
AIC	155.0152

TABLA 4: Interceptos del modelo de desempeño en inglés.

De la tabla (3) se puede afirmar que un aprendiz de sexo masculino con una edad promedio de 23 años tiene 1,54 veces más probabilidad de obtener un nivel de desempeño en inglés A2 que un aprendiz de sexo femenino.

Con relación a la disponibilidad del servicio de internet en la vivienda un aprendiz que cuente con este servicio tiene 1,13 veces más probabilidad de obtener un nivel de desempeño en inglés A2 que un aprendiz que no tenga internet.

Por otra parte, un aprendiz cuya madre tenga un nivel educativo de secundaria, tiene 1,35 veces más probabilidad obtener un nivel de desempeño en inglés A2 que un aprendiz cuya madre tenga nivel educativo primaria.

Modelo del desempeño en comunicación escrita

Variable	Estimación	Error estándar	Razón odds	Intervalo de confianza	
Edad	0.03640	0.03536	1.0371	0.9736	1.1047
Género [T.M]	0.17339	0.50749	1.1893	0.4431	3.1922
Educación de la madre [T.Secundaria]	0.02613	0.45146	1.0265	0.4220	2.4966
Educación del padre [T.Secundaria]	-0.18726	0.47149	0.8292	0.3282	2.0954
Internet [T.SI]	-0.39469	0.52245	0.6739	0.2342	1.9394

TABLA 5: Interpretación del modelo de desempeño en comunicación escrita.

	Value	Std. Error	t value
Nivel 1 Nivel 2	-1.2378	0.8616	-1.4366
Nivel 2 Nivel 3	1.2566	0.8462	1.4850
Nivel 3 Nivel 4	2.6878	0.8929	3.0102

Resid. Deviance	200.2853
AIC	216.2853

TABLA 6: Interceptos del modelo de comunicación escrita.

Respecto de la tabla (5) un aprendiz de género masculino tiene 1,05 veces más probabilidad de obtener un nivel 2 de desempeño en comunicación escrita respecto de un aprendiz de sexo femenino.

Un aprendiz cuya madre tenga un nivel educativo de secundaria, tiene 1,00 veces más probabilidad obtener un nivel de desempeño en nivel 2 que un aprendiz cuya madre tenga nivel educativo primaria.

Ahora, respecto a la disponibilidad de internet en la vivienda de los aprendices se observa que el Odds de bajo a alto nivel de comunicación escrita va disminuyendo para los aprendices que no cuentan con el servicio de internet respecto de quienes cuentan con este servicio.

Para evaluar la efectividad de los modelos se realizó una predicción de los resultados individuales de los 94 aprendices en desempeño de inglés y comunicación escrita. Para realizar esta predicción se tiene en cuenta que como resultado del uso de la RLO lo que se logra es una estimación de la probabilidad que el aprendiz obtenga cada uno de los posibles niveles (A1, A2, B1) y(Nivel 1, Nivel 2, Nivel 3, Nivel 4). Las predicciones se compararon con los resultados reales de los aprendices. El cálculo del porcentaje total de aciertos derivado de esta comparación constituye la base empleada para juzgar la efectividad del modelo. Este mismo criterio fue empleado por Rico, Hernández y Alonso (2014) donde los autores plantean que escogen como mejor solución aquella con la cual el porcentaje total de aciertos es máximo. Con los datos de los 94 aprendices se confeccionó una tabla donde se incluyó: las probabilidades estimadas, el pronóstico realizado, el resultado real y si hubo o no acierto. A manera de ejemplo, se expone en las tablas (7 y 8) el pronóstico realizado para tres de los 94 aprendices. Esta representa una muestra de la que fue confeccionada para todo el conjunto de aprendices.

Aprendiz	Prob de A1	Prob de A2	Prob de B1	Pronóstico	Nivel real	Acierto
Aprendiz X	0,34	0,52	0,14	A2	A1	NO
Aprendiz Y	0,30	0,54	0,16	A2	A1	NO
Aprendiz Z	0,41	0,48	0,1	A2	A2	SI

TABLA 7: Ejemplo de predicción para el desempeño en inglés

El aprendiz denotado como X (se emplea esta denominación para mantener el anonimato de los aprendices implicados) obtiene una mayor probabilidad en A2, es por eso que se pronostica un buen nivel. Como finalmente obtiene A1, se registra como acierto no logrado en el pronóstico. El mismo caso tenemos para el aprendiz Y, para el aprendiz Z se pronostica un nivel A2. como finalmente obtiene A2, se registra como acierto logrado en el pronóstico.

Aprendiz	Prob. de Nivel 1	Prob. de Nivel 2	Prob. de Nivel 3	Prob. de Nivel 4	Pronóstico	Nivel real	Acierto
Aprendiz X	0,43	0,47	0,07	0,03	Nivel 2	Nivel 2	SI
Aprendiz Y	0,52	0,41	0,05	0,02	Nivel 1	Nivel 3	NO
Aprendiz Z	0,29	0,54	0,12	0,05	Nivel 2	Nivel 2	SI

TABLA 8: Ejemplo de predicción para el desempeño en comunicación escrita

El aprendiz denotado como X obtiene una mayor probabilidad en Nivel 2, es por eso que se pronostica un bajo nivel. Como finalmente obtiene Nivel 2, se registra como acierto logrado en el pronóstico. para el aprendiz Y se pronostica un nivel 1. como finalmente obtiene Nivel 3, se registra como acierto no logrado en el pronóstico y para el aprendiz Z se pronostica un nivel 2. como finalmente obtiene Nivel 2, se registra como acierto logrado en el pronóstico. En la tabla (9) y (10) se resume los resultados del pronóstico ejecutado con los 94 aprendices.

Total Aprendices	94
Aciertos	45
Fallos	38
Celdas sin respuesta	11

TABLA 9: Resumen del pronóstico del desempeño en inglés

De los 94 aprendices cuyo pronóstico fue realizado, hubo 45 aciertos y 38 fallos, para un 47,87% de aciertos y un 40,43% de fallos. los restantes fueron celdas faltantes.

Total Aprendices	94
Aciertos	49
Fallos	34
Celdas sin respuesta	11

TABLA 10: Resumen del pronóstico del desempeño en comunicación escrita

De los 94 aprendices cuyo pronóstico fue realizado, hubo 49 aciertos y 34 fallos, para un 52,13% de aciertos y un 36,17% de fallos. los restantes fueron celdas faltantes.

Haciendo un análisis más detallado observamos que el modelo del desempeño en inglés tiene menor ajuste que el del desempeño en comunicación escrita. Pero aún así los dos presentan un bajo ajuste.

Calidad del ajuste de los modelos

	R^2 de Mc-Fadden	R^2 de Nagelkerke
Desempeño en inglés	0.1344141	0.2524223
Desempeño en comunicación escrita	0.1005516	0.2338349

TABLA 11: Calidad del ajuste de los modelos

Como los valores del R^2 de Mc-Fadden tanto para desempeño en inglés y comunicación escrita son menores que 0.2, podríamos decir que no se considera que los modelos presenten un buen ajuste.

Para los valores del R^2 de Nagelkerke para los dos desempeños nos da 0,2, podríamos decir que se considera una buena calidad de ajuste.

En general estos valores son cercanos a 0, lo cual nos indica que los modelos de desempeño en inglés y comunicación escrita no presentan un ajuste muy bueno.

5. Discusión

Dado el estudio descriptivo de los datos se tiene que la muestra corresponde a una mayoría de aprendices entre los 20 y los 28 años de edad (90 %), siendo la cantidad de mujeres superior al de los hombres, la educación de los padres en su mayoría es primaria, el (70 %) no cuenta con un servicio de internet en el hogar, los cuales un buen porcentaje (57 %) obtuvieron un nivel A1 en el desempeño en inglés y (53 %) obtuvieron nivel 2 (Mínimo) en el desempeño de comunicación escrita.

Se observa que el intervalo de confianza al 95 % para la razón de odds incluye al 1, esto permite asegurar, con una probabilidad menor que 0.05 de equivocación que las variables explicativas edad, género, educación de la madre, educación del padre y la disponibilidad de internet en la vivienda, no influyen de manera significativa en el desempeño de inglés y comunicación escrita.

Los resultados muestran que el modelo tiene una tasa de clasificación correcta del 47, 87 % una vez se empleó una estrategia simple de cotejo de los valores esperados y observados mediante simple inspección. El modelo indica que se deben considerar otras variables explicativas referenciadas en los estudios previos para cumplir el propósito de encontrar un modelo que pueda explicar los datos con un mínimo error y en este caso el ajuste del modelo no es apropiado.

CEDEAGARO debe contar con una mayor oferta de refuerzos académicos para el aprendizaje del idioma Inglés que permitan mejorar las competencias de los aprendices y obtener los más altos niveles de desempeño, así como el desarrollo de actividades académicas transversales (Club de conversación y lectura, laboratorios de idiomas, entre otros que logren motivar a los aprendices en el aprendizaje de esta lengua. Así como el desarrollo de estrategias que busquen mejorar el desempeño en comunicación escrita incentivando la lectura y la escritura, para este aspecto se puede a través del diseño de cursos empleando la plataforma de LMS

Referencias Bibliográficas

- Agresti, A. (2002), *Categorical data analysis*, John Wiley & Sons.
- Agresti, A. (2007), *An introduction to categorical data analysis*, Vol. 135, John Wiley & Sons.
- Agresti, A. (2013), *Categorical data analysis*, Vol. 482, John Wiley & Sons.
- Balaguer, E. P., Sinha, S. y Goitía, A. (2009), ‘Sobre la agrupación de niveles del factor explicativo en el modelo logit binario’, *Revista Colombiana de Estadística* **32**(2), 157–187.
- Díaz, L. G. y Morales, M. (2012), *Análisis estadístico de datos categóricos*, Universidad Nacional de Colombia.
- Duarte, B. R., Peña, W. Q. y Bustos, B. A. A. (2015), ‘Identificación de variables socioeconómicas e institucionales asociadas al desempeño académico de los aprendices sena en las pruebas saber pro 2012’.
- DUEÑAS, R. M. (2006), ‘Modelos de respuesta discreta en r y aplicación con datos reales’, *Universidad de Granada, España*.
- Gonzalez, J. J. y Cepeda, C. H. (2016), ‘Aplicación de un modelo de regresión logística con respuesta polinómica ordinal en el análisis del desempeño académico en matemáticas.’.
- Hosmer, D. W. y Lemeshow, S. (2000), *Applied logistic regression*, Vol. 398, John Wiley & Sons.

ICFES (2017), 'Icfes', url<http://www.icfes.gov.co>.

McCullagh, P. (1980), 'Regression models for ordinal data', *Journal of the royal statistical society. Series B (Methodological)* pp. 109–142.

Rico, J. J. H., Hernández, A. G. R. y Alonso, J. A. V. (2014), 'Empleo de la regresión logística ordinal para la predicción del rendimiento académico', *Investigación Operacional* **33**(3), 252–267.

Sánchez, Y. A. (2017), 'Análisis descriptivo de los resultados de las pruebas saber pro en la uptc del periodo 2012-2015', *PENSAMIENTO Y ACCIÓN* (22).



IDENTIFICACIÓN MORFOMÉTRICA DE ESCARABAJOS COPRÓFAGOS (Coleoptera: Scarabaeidae: Scarabaeinae) UN EJERCICIO DE APROXIMACIÓN A LA TAXONOMÍA DESDE LA ESTADÍSTICA

Especialización en Estadística

OSCAR FELIPE MORENO MANCILLA^{1,a}, JOSÉ DANIEL CARO CASTILLO^{1,b}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

Debido a la necesidad de identificar correctamente escarabajos coprófagos de una forma rápida y de bajo costo, se propone el desarrollo de una herramienta basada en medidas morfométricas y técnicas estadísticas multivariadas para la determinación taxonómica. Se generó una base de datos a partir de 60 individuos pertenecientes a tres especies, a los cuales se les midieron un total de 10 variables. Se realizaron análisis estadísticos descriptivos, análisis de componentes principales y se propusieron análisis discriminantes lineales y cuadráticos como mecanismo de clasificación supervisado que posteriormente fueron evaluados mediante validación cruzada. Se encontró que tanto el análisis discriminante lineal como el cuadrático son idóneos para clasificar éstos organismos en las tres especies utilizadas, con una tasa de buena clasificación del 100 %. Sin embargo esto se debe a que estas tres especies fueron seleccionadas por sus marcadas diferencias en tamaño con fines demostrativos, y es posible que estos análisis discriminantes no sean tan efectivos para clasificar individuos entre grupos de especies que compartan similitudes en magnitud y variación de sus atributos morfométricos. Se recomienda ampliar la base de datos y realizar pilotajes entre especies similares para poder encontrar qué variables morfométricas son las más adecuadas para su clasificación.

Palabras clave: Identificación taxonómica, morfometría, estadística multivariada, escarabajos coprófagos..

Abstract

Due to the necessity of dung beetles correct identification in a fast and low budget way, here, a tool development based on morphometric measures and multivariate statistical techniques is proposed for the taxonomic determination. A database was generated from 60 individuals belonging to three different species, where a total of ten variables were measured. There were made descriptive statistical analysis, principal components analysis and there were proposed lineal and quadratic discriminant analysis as a mechanism of supervised classification that subsequently were tested by means of a crossed validation. There was found that as the lineal discriminant analysis as well as the quadratic one were suitable to classify these organisms in the three species used, with a 100 % rate of good classification. However this is due to the fact that these species were selected by their highlighted differences in size with demonstrative purposes, and it is possible that these discriminant analyses aren't so effective to classify individuals between groups of species who share magnitude and variation similarities of their morphometric qualities. It is recommended to extend the database and to make tests between similar species in order to find what morphometric variables are more suitable and convenient for their classification.

Key words: Taxonomic identification, morphometry, multivariate statistics, dung beetles..

^aEstudiante de Especialización en Estadística. E-mail: felipemancilla@gmail.com

^bEstudiante de Especialización en Estadística. E-mail: dcarodaniel91@gmail.com

1. Introducción

La comparación morfológica basada en la presencia o ausencia de estructuras ha sido el eje fundamental de la taxonomía biológica a lo largo de su historia (Benítez y Püschel 2014), generando un panorama en el que conocer un sinnúmero de nombres asociados a cada estructura corporal es necesario para lograr la correcta identificación de una entidad biológica. De esta manera, la taxonomía se convierte en una ciencia altamente especializada de acceso restringido para otras ramas de conocimiento como la Ecología, la Biología de la Conservación, la Ecología de la Restauración o la Entomología Forense entre otras, que precisan de métodos de identificación rápidos y de bajo costo (Vásquez y Liria 2012).

En particular, la taxonomía de escarabajos coprófagos (Coleoptera: Scarabaeidae: Scarabaeinae) ha tenido un enfoque tradicional netamente descriptivo, y a pesar de ser uno de los grupos de insectos mejor estudiados, aún presenta grandes vacíos a nivel taxonómico y sistemático (Simmons y Ridsdill-Smith 2011). Además, aunque la identificación hasta género de estos insectos es posible mediante diversas claves taxonómicas (Medina y Toro 2000, Vaz-de Mello, Edmonds, Ocampo y Schoolmeesters 2011), esta, no deja de ser una labor que exige en gran medida un conocimiento profundo de la anatomía y morfología de estos animales; a lo que se suma que la identificación hasta nivel de especie, en muchos casos requiere claves especializadas, revisiones para los géneros, disección y comparación morfológica de genitales, y conocimiento de distribución biogeográfica de las especies. Si se tiene en cuenta el enorme potencial que tienen los escarabajos estercoleros como grupo indicador de la calidad de un hábitat o de procesos de restauración, debido a su participación en innumerables procesos y servicios ecosistémicos (Nichols, Spector, Louzada, Larsen, Amezquita, Favila et al. 2008), lo anterior, podría llegar a ser una limitación en cuanto a que una mala identificación por parte de un investigador no especializado en su taxonomía, podría generar información errónea.

En este sentido, la morfometría como una herramienta matemática con fundamentos biológicos que permite caracterizar estadísticamente los atributos medibles de la “forma” y el “tamaño” de un organismo, aprovechando al máximo la información geométrica que poseen sus estructuras (Toro Ibacache, Manriquez Soto y Suazo Galdames 2010), podría ser de utilidad en el desarrollo de sistemas de identificación que integren la taxonomía tradicional de enfoque descriptivo con atributos estadísticos de enfoque cuantitativo permitiendo la determinación fácil, rápida y correcta de organismos a investigadores no taxónomos (Guisande, Manjarrés-Hernández, Pelayo-Villamil, Granado-Lorencio, Riveiro, Acuña, Prieto-Piraquive, Janeiro, Matías, Patti et al. 2010).

La morfometría ha sido desarrollada después de la segunda mitad del siglo XX de la mano con técnicas estadísticas multivariadas que proporcionan análisis robustos y herramientas gráficas para la cuantificación y visualización de la variación morfológica (Jaramillo, 2017). En contextos biológicos ha sido usada en ramas como la Ecología y la Biología Evolutiva (Fontanillas Pérez 2007, Prieto y De Polanco 2009, Sánchez 2014, Vásquez y Liria 2012), mientras que a nivel taxonómico ha sido utilizada en estudios de caso para separar especies mediante diferencias estadísticas en la variabilidad de sus estructuras (Vásquez y Liria 2012).

Sin embargo, en el caso específico del desarrollo de herramientas computacionales que permitan la identificación de individuos, ha sido poco explorada siendo el único caso el software iPez que permite la determinación taxonómica de peces adultos (Guisande, Manjarrés-Hernández, Pelayo-Villamil, Granado-Lorencio, Riveiro, Acuña, Prieto-Piraquive, Janeiro, Matías, Patti et al. 2010). Por lo tanto el objetivo del presente estudio es proponer mediante un ejercicio de aproximación, el desarrollo de una herramienta de identificación morfométrica de escarabajos coprófagos, que de la mano de la taxonomía clásica, facilite a investigadores no taxónomos la correcta identificación de éstos insectos de alto interés ecológico.

2. Metodología

2.1. Recolección de los datos

Se consolidó una base de datos que contiene la medición de 10 variables morfométricas a 20 individuos de las especies *Canthidium sp 1*, *Canthon aff politus* y *Eurysternus caribaeus* respectivamente, para un total

de 60 escarabajos medidos y 600 datos (Tabla 1, Figura 1, Anexo 1.); todas las variables fueron medidas en milímetros. Estos escarabajos fueron seleccionados aleatoriamente de la “Colección entomológica del Museo de Historia Natural Luis Gonzalo Andrade” “Universidad Pedagógica y Tecnológica de Colombia UPTC”, de la “Colección de Escarabajos Coprófagos de Colombia” del Instituto Alexander von Humboldt y de la “Colección de Referencia de Escarabajos Coprófagos convenio Ecopetrol-UPTC”. Las mediciones fueron realizadas sobre una fotografía dorsal de cada individuo tomada con un microscopio estereoscópico Leica EZ4 $\frac{W}{E}$, equipado con cámara digital de alta resolución Leica ICC50 $\frac{W}{E}$ y el software de procesamiento de imagen Leica Application Suit EZ LAS EZ, que permiten la captura de imágenes y posterior medición de alta precisión.

Variable	Abreviación	Medida
Longitud total	LT	Máxima distancia entre el margen anterior de la cabeza y el margen posterior de los élitros
Longitud de la cabeza	LC	Máxima distancia entre el margen anterior y posterior de la cabeza
Longitud del pronoto	LP	Máxima distancia entre el margen anterior y posterior de la cabeza
Longitud de los élitros	LE	Máxima distancia entre el margen anterior y posterior de la cabeza
Ancho de la cabeza	AC	Máxima distancia entre el margen derecho e izquierdo de la cabeza
Ancho del pronoto	AP	Máxima distancia entre el margen derecho e izquierdo del pronoto
Ancho de los élitros	AE	Máxima distancia entre el margen derecho e izquierdo de los élitros
Longitud del fémur anterior	LFA	Distancia entre el margen apical y distal del fémur anterior
Longitud del fémur posterior	LFP	Distancia entre el margen apical y distal del fémur posterior
Ancho del fémur posterior	AFP	Distancia entre el margen izquierdo y derecho de la zona distal del fémur posterior

TABLA 1: Medidas morfométricas realizadas a los escarabajos.

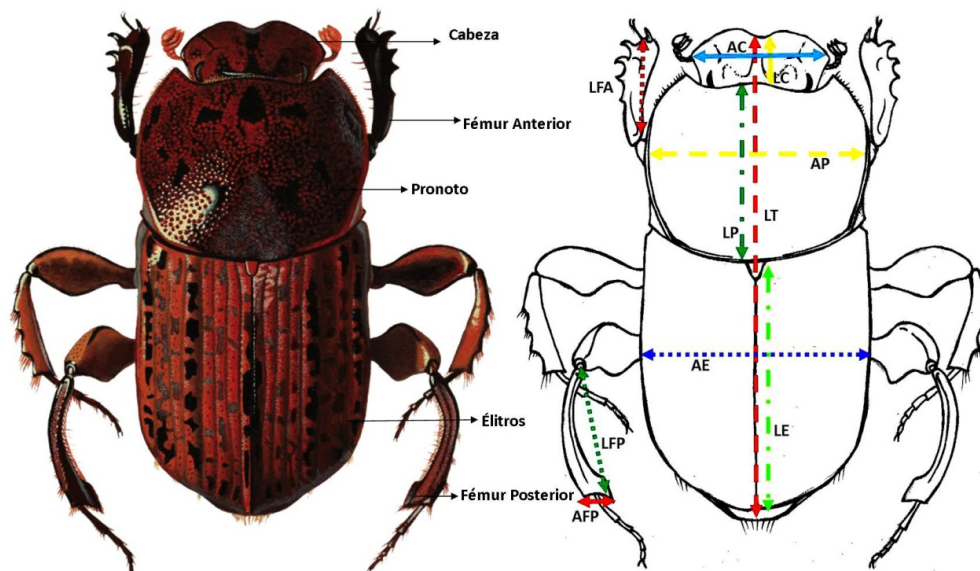


FIGURA 1: Partes de interés en este estudio de *Eurysternus marmoreus*, ubicación de las variables morfométricas medidas. Tomada y modificada de Genier (1996).

2.1.1. Análisis de los datos

Caracterización morfométrica

Se realizó un análisis descriptivo de los datos calculando la media y la desviación estándar de cada variable medida, para cada especie, con el fin de realizar una primera exploración de los datos. Para éste análisis se utilizó la totalidad de los individuos medidos.

Análisis de componentes principales

Se realizó un análisis en componentes principales (PCA por sus siglas en Inglés Principal Component Analysis) previa normalización de los datos, con el fin de explorarlos de una manera multivariada de tal manera que sea posible encontrar una combinación lineal de las variables, que expliquen su variación sin pérdida de información. El ACP es una técnica de análisis estadístico multivariante que reduce la dimensión persiguiendo obtener un menor número de variables, combinación lineal de las primitivas, que se denominan componentes principales, cuya posterior interpretación permitirá un análisis más simple del problema estudiado. Como medida de la cantidad de información incorporada en una componente se utiliza su varianza. Es decir, cuanto mayor sea su varianza mayor es la información que lleva incorporada dicha componente. Por esta razón se selecciona como primera componente aquélla que tenga mayor varianza, mientras que, por el contrario, la última es la de menor varianza (Díaz 2002, Molina y García 2006).

Análisis discriminante

Se propusieron como métodos de clasificación supervisada un análisis discriminante lineal (LDA por sus siglas en Inglés Linear Discriminant Analysis) y un análisis discriminante cuadrático (QDA por sus siglas en Inglés Quadratic Discriminant Analysis). Estos, son métodos de clasificación supervisados de variables cualitativas en el que dos o más grupos son conocidos a priori y nuevas observaciones se clasifican en uno de ellos en función de sus características. Haciendo uso del teorema de Bayes, LDA y QDA estiman la probabi-

lidad de que una observación, dado un determinado valor de los predictores, pertenezcan a cada una de las clases de la variable cualitativa, $P(Y=k|X=x)$. Finalmente se asigna la observación a la clase k para la que la probabilidad predicha es mayor (Diaz 2002, Molina y García 2006).

Estos análisis requieren un conjunto de datos de entrenamiento (training data) en el que se conoce a que grupo pertenece cada observación, por lo cual, fueron seleccionados aleatoriamente 4 individuos de cada especie, para un conjunto de datos de entrenamiento de 12 individuos.

A continuación se calcularon las probabilidades previas (prior probabilities) como la proporción esperada de observaciones que pertenecen a cada grupo; se determinó la homogeneidad de varianzas; y por último se calculó el resultado de la función discriminante cuyo resultado determina a qué grupo se asigna cada observación. Finalmente se realiza el cálculo de la tasa de buena clasificación mediante una prueba de validación cruzada usando el conjunto de datos de entrenamiento.

Todos los análisis estadísticos y salidas gráficas fueron realizados en el paquete estadístico R versión 3.4.1, interfaz RStudio versión 1.1.453 (R Core Team 2017).

Resultados

Caracterización morfométrica

Variable	<i>Canthidium sp1</i>	<i>Canthon politus</i>	<i>Eurysternus marmoreus</i>
LT	5.3	8.8	11.71
LC	0.75	1.22	0.9
LP	1.65	2.53	4.36
LE	2.53	4.69	6.38
AC	1.77	2.64	3.44
AP	0.55	0.93	1.11
AE	3.49	5.46	5.64
LFA	1.41	2.37	2.49
LFP	1.57	3.16	3.53
AFP	0.63	0.57	0.73

TABLA 2: Promedio para las diferentes variables de cada especie.

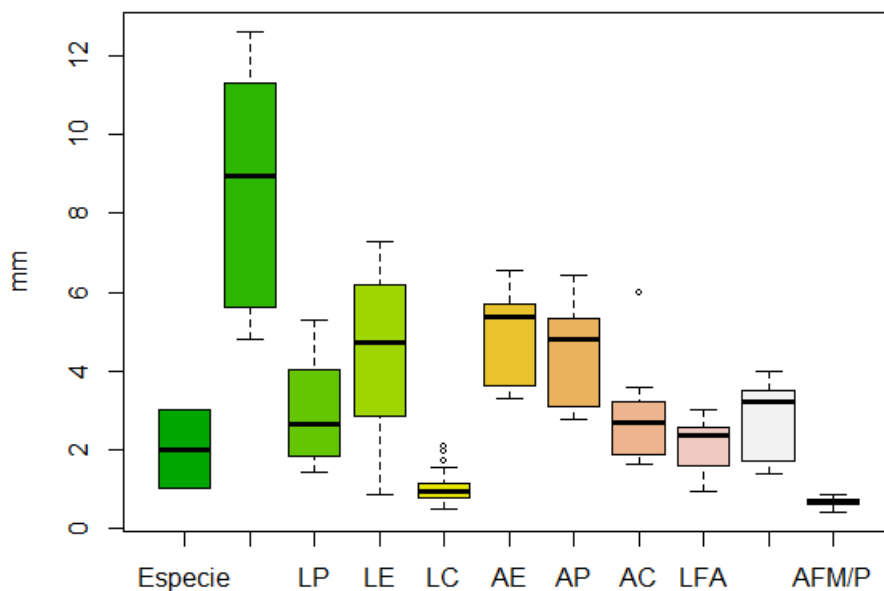
Boxplot de las variables morfométricas de escarabajos

FIGURA 2: Diagrama de cajas para las variables medidas en las tres especies.

Análisis de componentes principales

EL análisis exploratorio de los datos permitió revisar a grandes rasgos características principales en los datos, de esta manera se observó, que la correlación entre las variables es diferente de cero ($cor \neq 0$), lo que indica que para el caso es acertado realizar un análisis de reducción de dimensión a través de componentes principales. Además, se detectaron posibles de datos atípicos, probablemente debido a que la magnitud en las medidas para *Eurysternus marmoreus* sobrepasan las mediciones de las especies restantes. Sin embargo con la normalización de los datos se evitan problemas de inflación de varianza que puedan ser causados por estos datos.

Se puede observar (Tabla 3.), que la componente principal 1 (CP1) recoge mayor información para la variable AP (ancho del pronoto), mientras que la variable ancho del fémur posterior (AFM/P) en esta componente, es muy débil y probablemente poco informativa. Por otro lado, la componente principal 2 (CP2) recoge menos información que CP1, mientras que la componente principal 3 (CP3) explica menos información que CP1 Y CP2, por lo cual las tres primeras componentes bastarían para explicar la variabilidad de las medidas morfométricas de las tres especies de escarabajos.

El porcentaje de varianza explicada por CP1 es del 73.88% siendo la componente principal que explica la mayor variabilidad de los datos, seguida de CP2 explicando el 12,31% de la variabilidad que no cubre la primera componente CP1. Nótese entonces que entre mayor sea el número de componentes formuladas, menor proporción de varianza se explica por las componentes.

Debido a que la transformación ortogonal se realiza a partir de la matriz de correlación, la varianza de cada una de las variables originales debe ser uno, por lo tanto, la varianza de cada componente principales no debe ser menor que uno (Kaiser, 1960). De ésta forma se concluye que la variabilidad total de los datos originales es explicada por las dos primeras componentes CP1, CP2 con una proporción acumulada de varianza explicada del 86.19%.

Al graficar los componentes principales junto con los datos (Figura 3.), se observan tres grupos bien diferenciados, por lo que se espera que los análisis discriminantes sean una buena alternativa para la clasificación de individuos a pesar de que *Canthidium sp 1* no se explica alrededor de las variables de CP1

	CP1	CP 2	CP 3	CP 4	CP 5	CP 6	CP 7	CP 8	CP 9	CP 10
Desviación estándar	2718	1109	0.8782	0.4752	0.4277	0.2928	0.2267	192	0.1194	109
Proporción Varianza	0.7388	123	0.0771	0.0225	0.0182	0.0085	0.0051	3	0.0014	1
Proporción acumulada	0.7388	861	939	961	979	988	993	997	998	1.0000

TABLA 3: Porcentaje de Variabilidad explicada por los componentes principales.

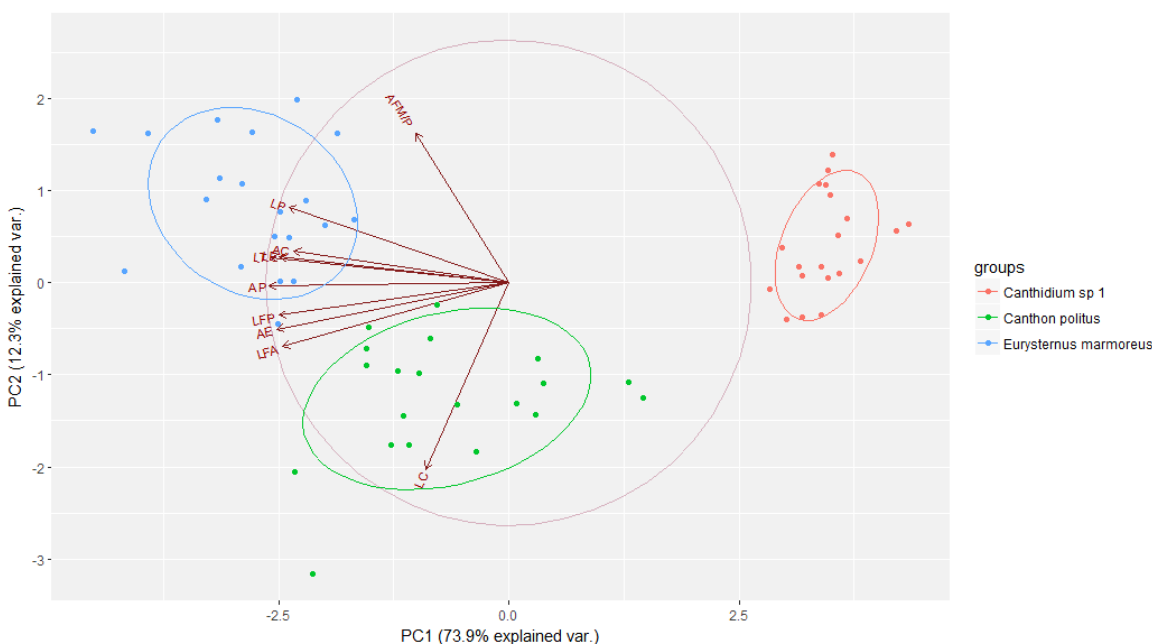


FIGURA 3: Diagrama de componentes principales para las tres especies.

Análisis discriminante

Análisis discriminante lineal LDA: Se puede observar que las mejores variables clasificadoras para las tres especies de escarabajos son LT, LP y LE debido a que sus distribuciones presentan diferenciación y poco solapamiento, al contrario de las demás variables que debido a sus fuertes solapamientos no podrían ser usadas en la clasificación de individuos en estas tres especies. De manera intuitiva se evidencia la falta de normalidad univariante en cada una de las variables (Figura 4.).

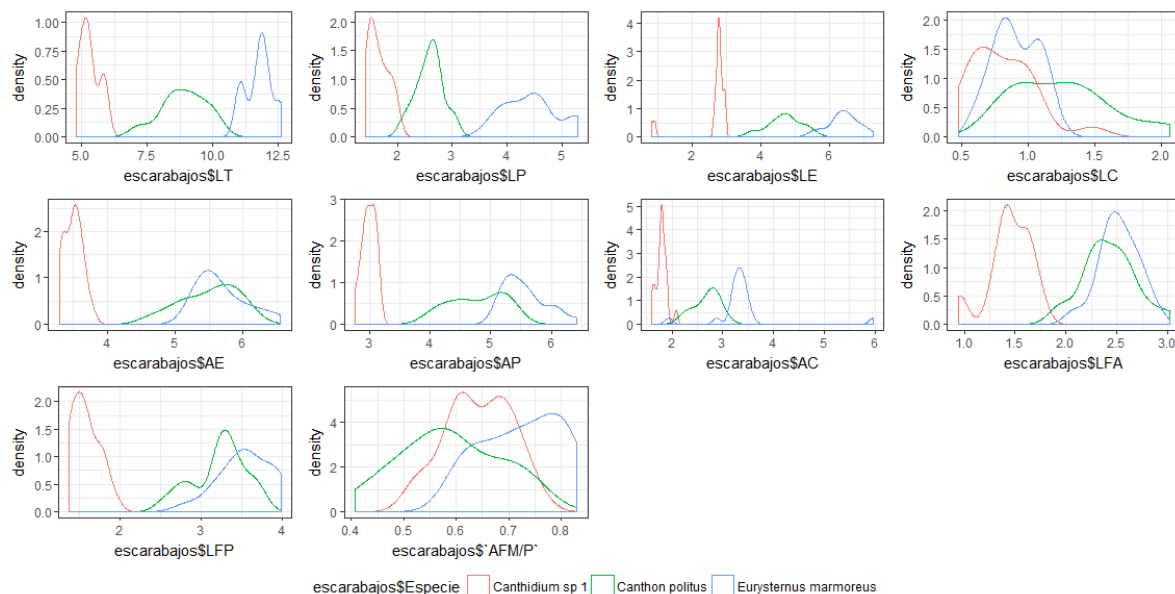


FIGURA 4: Diagrama de densidad de distribución para cada variable.

El test de Anderson Darling evidencia que para *Canthidium sp 1*, no hay presencia de normalidad en la variable LE, mientras que en *C. politus*, los datos para todas variables, tienen distribución normal, y a su vez *E. marmoreus*, no hay tiene normalidad en los datos registrados para la variable AC.

De igual manera, el test de normalidad multivariante de Henze-Zirkler, garantiza distribución normal multivariante para la especie *E. marmoreus*. Sin embargo, este resultado no es importante, pues si bien la presencia de normalidad en la distribución de los datos en una condición, no es necesaria, debido a la robustez que tiene la técnica del análisis discriminante le permite ser aplicada haciendo caso omiso de los supuestos de normalidad pero arriesgando un poco la precisión.

El Test de Bartlett para probar la homogeneidad de varianzas arroja un p-valor de $2.2e-16$ lo que a una confianza del 95 % garantiza varianza constante entre las tres especies, por lo tanto es válido ejecutar un LDA.

El LDA arroja que las probabilidades a priori de que una observación sea seleccionada en cada una de las especies son de 0.29 para *Canthidium sp1*, 0.33 para *C. politus* y para *E. marmoreus* 0.37. Por otro lado, el primer discriminante lineal (LD1) tiene un grado de importancia de 84 % mientras que el segundo discriminante lineal (LD2) es de tan solo el 16 %. Además se encuentra el promedio de cada predictor dentro de cada clase, usados por el modelo como estimadores de μ_k (Tabla 4).

Los coeficientes para LD1 y LD2 tienen un mayor grado de importancia relativa para las variables AE, AC, AFM /P indicando una fuerte relación entre las funciones y estas variables.

Coeficientes de discriminantes lineales		
	LD1	LD2
LT	1.5732914	-0.39447143
LP	1.3968347	1.75884640
LE	0.3555113	0.54117811
LC	-2.5779326	-1.34964434
AE	-4.0042434	-4.81850416
AP	2.6382716	0.94903279
AC	-0.3724011	0.42381081
LFA	1.3196793	0.64317309
LFP	1.4801883	0.08069198
Grado de importancia		
	0.8393	0.1607

TABLA 4: Coeficientes de discriminantes lineales y porcentaje de importancia.

Como se puede observar en la Figura 5, el LDA discrimina claramente los datos en tres grupos correspondientes a las especies, indicando que es un análisis que puede servir para la clasificación de individuos en éstas tres especies de escarabajos.

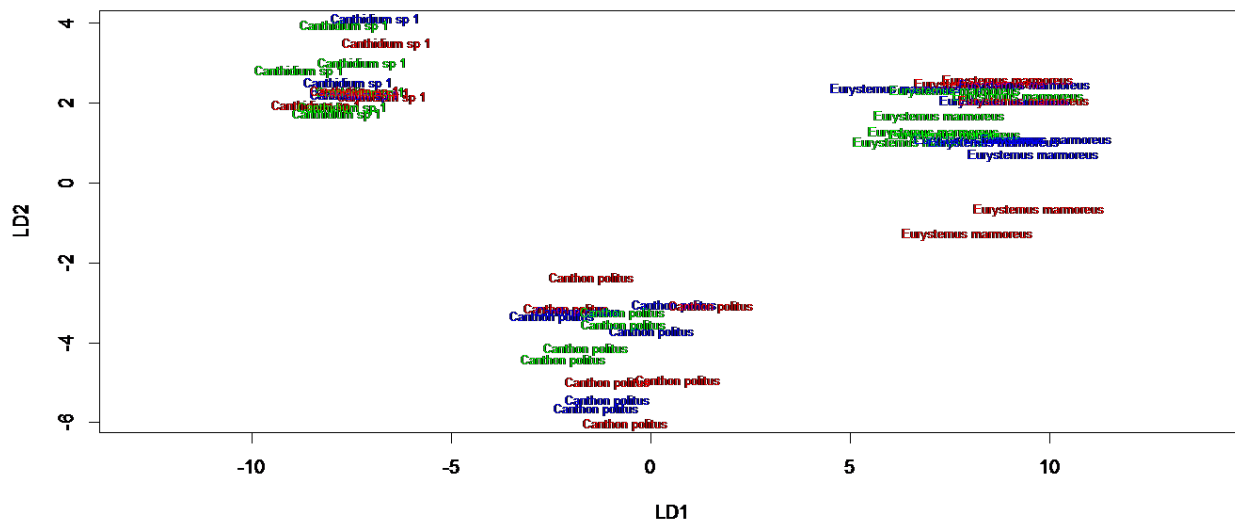


FIGURA 5: Diagrama de coeficientes modelo discriminante lineal para las tres especies.

Modelo discriminante cuadrático QDA: Este modelo arroja unas probabilidades a priori exactamente iguales a las mostradas por LDA, sugiriendo resultados similares para ambos procesos de clasificación. Así mismo las medias calculadas para cada grupo presentan los mismos valores.

Validación cruzada: Arroja una tasa de buena clasificación para ambos modelos de 100% y una precisión del 100% indicando que no durante la prueba, todos los individuos del set de datos de entrenamiento fueron

clasificados correctamente en la especie a la que pertenecía. Sin embargo, estos valores deben ser validados con más cantidad de datos pues se espera siempre un sesgo debido a la magnitud de los datos.

Clase predicha	Clase Real		
	<i>Canthidium sp 1</i>	<i>Canthon politus</i>	<i>Eurysternus marmoreus</i>
<i>Canthidium sp 1</i>	6	0	0
<i>Canthon politus</i>	0	4	0
<i>Eurysternus marmoreus</i>	0	0	2

TABLA 5: Proporción de individuos correctamente clasificados

3. Discusión

Las variables seleccionadas son una buena fuente de variación morfométrica de utilidad en el posible desarrollo de un sistema de clasificación taxonómico basado en caracteres cuantitativos. Atributos como la longitud total (LT) como medida directa del tamaño de un escarabajo coprófago, o la longitud del pronoto (LP), relacionada indirectamente con la capacidad de vuelo (Hernández et al., 2011) representan información con potencial de clasificación por lo menos a nivel de género. Al contrario de lo que se podría esperar, variables como longitud del fémur posterior (LFP) y ancho del fémur (AFP) posterior, no resultaron informativas a pesar de ser atributos estrechamente relacionados con hábitos ecológicos que diferencian grupos taxonómicos de escarabajos como lo son el hecho de rodar esferas de estiércol (escarabajos rodadores), cavar túneles para luego incorporar estiércol (Escarabajos cavadores), o simplemente no transportarlo (Escarabajos residentes) (Cultid, Medina, Martínez, Escobar, Constantino y Betancur 2012).

De igual forma, los análisis discriminantes, tanto lineal como cuadrático tuvieron el máximo de rendimiento, indicando que pueden ser herramientas estadísticas robustas con potencial para clasificación de individuos.

Sin embargo, vale la pena notar que este análisis representa una pequeña aproximación a la aplicación de la estadística en la taxonomía, y que las especies incluidas en el análisis fueron seleccionadas deliberadamente por sus evidentes diferencias en la magnitud de su tamaño, lo cual pone en duda la alta efectividad que tuvieron los métodos de análisis discriminante propuestos. Estas tres especies pertenecen a tres géneros de escarabajos coprófagos muy diferentes en cuanto a morfología, ecología e historias evolutivas por lo que era de esperarse un buen funcionamiento de los mecanismos de clasificación.

De igual forma, existen géneros de escarabajos coprófagos cuyas especies presentan una mínima variabilidad en el tamaño de sus estructuras, tal es el caso del género *Scybalocanton*, *Onthophagus* u *Uroxys*. Esto, podría dificultar el uso de éstas técnicas por lo que es necesario un estudio de caso para encontrar cual es la variable más informativa dentro de cada género, y así afinar los análisis clasificadores de acuerdo a cada caso particular.

Se recomienda tener un mayor número de escarabajos medidos con el fin de hacer más robustos los análisis, al igual que incluir en la base de datos información con respecto a la procedencia de cada individuo, con el fin de detectar posibles variaciones entre poblaciones producidas por presiones ambientales relacionadas a los ecosistemas que una vez habitaron.

La consolidación de una base de datos con la información necesaria para desarrollar una herramienta compu-

tacional que permita la identificación de escarabajos coprófagos a diferentes niveles taxonómicas, con seguridad podrá arrojar luz sobre otros aspectos de la biología de éstos insectos, como su comportamiento, la funcionalidad de sus atributos, la evolución de su forma y tamaño, su relación con los ecosistemas, su distribución, la diversificación de sus especies entre muchas otros.

Referencias Bibliográficas

- Benítez, H. A. y Püschel, T. A. (2014), ‘Modelando la varianza de la forma: morfometría geométrica aplicaciones en biología evolutiva’, *International Journal of Morphology* **32**(3), 998–1008.
- Cultid, C., Medina, C., Martínez, B., Escobar, A., Constantino, L. y Betancur, N. (2012), ‘Escarabajos coprófagos (scarabaeinae) del eje cafetero: guía para el estudio ecológico’, *WCS-Colombia, CENICAFÉ y Federación Nacional de Cafeteros. Villa María. Colombia* .
- Díaz, L. (2002), *Estadística multivariada: inferencia y métodos*, Bogotá: Universidad Nacional de Colombia.
- Fontanillas Pérez, J. C. G.-C. A. (2007), ‘Diferencias morfométricas entre la abeja apis mellifera iberica y la abeja rusa de la región de primorsky; morphometrics differences between the bee apis mellifera iberica and the russian bee of the primorsky region’, *Revista Complutense de Ciencias Veterinarias* **1**(1), 8–19.
- Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J., Patti, C. et al. (2010), ‘Ipez: an expert system for the taxonomic identification of fishes based on machine learning techniques’, *Fisheries Research* **102**(3), 240–247.
- Hernández, M. I., Monteiro, L. R. y Favila, M. E. (2011), ‘The role of body size and shape in understanding competitive interactions within a community of neotropical dung beetles’, *Journal of Insect Science* **11**(1), 13.
- Jaramillo, N. (2011), ‘Morfometría geométrica: Principios teóricos y métodos de empleo’, *Research Gate* **7**(4), 23.
- López, J. M. y Herrero, J. G. (2006), ‘Técnicas de análisis de datos’, *Aplicaciones Prácticas utilizando Microsoft Excel y WEKA* .
- Mancilla, O. F. M. y Rendón, F. M. (2016), ‘Variación en las abundancias de homocopris achamas (harold, 1867)(coleoptera: Scarabaeidae: Scarabaeinae) en el páramo de rabanal, boyacá-colombia/variation in the abundances of homocopris achamas (harold, 1867)(coleoptera: Scarabaeidae: Scarabaeinae) in the rabanal paramo, boyacá-colombia’, *CIENCIA EN DESARROLLO* **7**(2), 67–74.
- Medina, C. A. y Toro, A. L. (2000), ‘Clave ilustrada para la identificación de géneros de escarabajos coprófagos (Coleoptera: Scarabaeinae) de Colombia’, *Caldasia* **22**(2), 209–315.
- Molina, J. y García, J. (2006), ‘Técnicas de análisis de datos: aplicaciones prácticas utilizando Microsoft Excel y Weka’, *Universidad Carlos III de Madrid España* .
- Nichols, E., Spector, S., Louzada, J., Larsen, T., Amezquita, S., Favila, M. et al. (2008), ‘Ecological functions and ecosystem services provided by scarabaeinae dung beetles’, *Biological conservation* **141**(6), 1461–1474.
- Prieto, R. S. y De Polanco, M. M. E. (2009), ‘Análisis morfométrico de la pigmentación en especies colombianas de drosophila (diptera: Drosophilidae) del grupo repleta/morphometric analysis of the pigmentation in colombian species of drosophila (diptera: Drosophilidae) from the repleta group’, *Revista Colombiana de Entomología* **35**(2), 224.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<https://www.R-project.org/>

- Sánchez, M. V. (2014), Morfometría geométrica aplicada al estudio evolutivo de los espermatozoides y su relación con determinantes de la fertilidad en roedores, PhD thesis, Universidad Autónoma de Madrid.
- Simmons, L. W. y Ridsdill-Smith, T. J. (2011), *Ecology and evolution of dung beetles*, John Wiley & Sons.
- Toro Ibacache, M. V., Manriquez Soto, G. y Suazo Galdames, I. (2010), 'Morfometría geométrica y el estudio de las formas biológicas: de la morfología descriptiva a la morfología cuantitativa', *International Journal of Morphology* **28**(4), 977-990.
- Vásquez, M. y Liria, J. (2012), 'Morfometría geométrica alar para la identificación de *chrysomya albiceps* y *c. megacephala* (diptera: Calliphoridae) de venezuela', *Revista de Biología Tropical* **60**(3).
- Vaz-de Mello, F. Z., Edmonds, W., Ocampo, F. C. y Schoolmeesters, P. (2011), 'A multilingual key to the genera and subgenera of the subfamily scarabaeinae of the new world (coleoptera: Scarabaeidae)', *Zootaxa* **2854**(1), 1-73.



FORMULACIÓN Y VALIDACIÓN DE CUESTIONARIO PARA MEDICIÓN DE LAS CONDICIONES SOCIOECONÓMICAS DE LOS TRABAJADORES FABRILES DE LAS ACERÍAS DEL DEPARTAMENTO DE BOYACÁ

Especialización en Estadística

SEBASTIAN JOYA TÉLLEZ^{1,a}, EDGAR GABRIEL DÍAZ MELGAREJO^{1,b}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

El artículo tiene por objeto mostrar el proceso de construcción metodológica y validación de un instrumento estadístico destinado a la medición de las condiciones socioeconómicas y socio laborales de los trabajadores fabriles del acero de las tres productoras férricas más importantes del Departamento de Boyacá. El documento está compuesto por cuatro partes: En primer lugar, se desarrolla una introducción en la cual se esboza la importancia de la comprensión de los procesos continuados de tercerización y precarización de los trabajadores industriales en la actualidad; en segundo lugar, se realiza un recuento del conjunto de teorías e investigaciones que soportan la necesidad de esta discusión, haciendo especial énfasis en las teorías sociológicas contemporáneas que intenta desvelar la complejidad del cambio de la estructura del régimen de acumulación fordista y del mundo del trabajo; en tercer lugar, se desarrolla la exposición y explicación del diseño y aplicación de los procedimientos de validación y medición de la consistencia interna del cuestionario, poniendo especial énfasis en la utilización del Alfa de Cronbach y en la proyección del marco y el diseño muestral para un tentativo desarrollo de proyecto; por último, se presentan los resultados encontrados en clave de la consistencia interna del test y de la necesidad de reajuste para poder ser aplicada a la población de trabajadores fabriles de las empresas acereras del corredor Industrial del Departamento de Boyacá.

Palabras clave: Trabajadores, Precarización, Industria, Validación, Fiabilidad, Instrumento de Medición.

Abstract

The purpose of this paper is showing the process of methodological construction and validation of a statistical instrument aimed at measuring the socioeconomic and socio-labor conditions of the steel workers of the three most important ferrous producers in Boyacá State area. The document is composed of four parts: First, an introduction that is developed which outlines the importance of understanding the continuous processes of outsourcing and precariousness of industrial workers in our territories. Secondly, a recount is made of the set of theories and research that support the need for this discussion, with special emphasis on contemporary sociological theories that try to reveal the complexity of the change of the structure of the Fordist accumulation regime and the world of work. Thirdly, the presentation and explanation of the design and application of the validation and measurement procedures of the internal consistency of the questionnaire is developed, placing special emphasis on the use of Cronbach's alpha and

^aEstudiante de Especialización en Estadística. E-mail: Sebastian.joya@uptc.edu.co

^bEstudiante de Especialización en Estadística. E-mail: Edgargabriel.diaz@uptc.edu.co

the projection of the frame and sample design for a tentative project development. Finally, we present the results found in terms of the internal consistency of the test and the need for readjustment to be applied to the population of factory workers of the steel companies of the industrial corridor of the Boyacá state area.

Key words: Workers, Precarization, Industry, Validation, Reliability, Measurement Instrument..

1. Introducción

La población de trabajadores y trabajadoras del acero del “corredor industrial” del departamento de Boyacá, en el transcurso de los últimos 25 años, ha sido sujeta a una transformación progresiva de las formas de vinculación y contratación laboral, a partir de la creación de la ley 50 de 1990, la implementación de la subcontratación por medio de la ley 789 de 2002 y la incursión de la llamadas cooperativas de trabajo asociado en la primera década del siglo XXI. Este fenómeno ha cambiado la realidad social de la región central en la medida que ha generado una suerte de diferenciación entre los nuevos trabajadores de subcontrata y los antiguos ligados a un trabajo “normal”.

De esta manera, se ha venido disminuyendo en términos globales y de manera paulatina el número de trabajadores, mediante despidos selectivos y masivos, así como el número de trabajadores vinculados por contrato a término indefinido dentro de las industrias. Este cambio no ocurre únicamente con el conjunto de tareas no misionales para el mantenimiento de las empresas, también se encuentra vinculado al proceso mismo de producción, factor que va en contravía de la normatividad vigente en tanto que es considerado como falsa tercerización. Este fenómeno excede las tipificaciones teóricas realizadas hasta el momento sobre la subcontratación, en la medida que rompe con la dualización¹ del trabajo periférico y con el sostenimiento del empate técnico el subcontrata y el trabajo especializado tipificado como directo, planteándose un proceso de homogenización tendiente hacia la precarización (Bouffartigue y Busso 2010) y el establecimiento de la subcontratación como política empresarial única.

Es por esto que es de vital importancia realizar una lectura coherente y profunda sobre de lo que ha venido sucediendo con el modelo productivo regional y con los trabajadores que históricamente le han compuesto. Es menester de la academia plantear nuevas formas de interpretación que permitan auscultar la realidad que viven los asalariados así como desarrollar instrumentos que permitan medir el impacto de la aplicación de los diferentes tipos de contratación tipificados en la normatividad colombiana, en la configuración de las condiciones socioeconómicas de los trabajadores fabriles de las empresas productoras del acero en Boyacá (Acerías Paz de Rio, Diaco-Gerdau, Sidenal).

2. Referente Conceptual

Los análisis en torno a la situación de los trabajadores y la forma estos como se insertan en el modelo productivo no son nuevos en la órbita académica. Desde principios del siglo XIX se ha visto un constante debate sobre el papel fundamental que juega el trabajo en la producción del valor y en la creación de la riqueza, así como en los efectos que el asalariamiento tiene en la fabricación de las condiciones de vida de los actores productivos. Este debate sostenido por los principales exponentes de la economía política (David Ricardo, Adam Smith, Jean-Baptiste Say) y profundizado en su extensa obra por (Marx 2008), puso de presente el rol clave que juegan los trabajadores asalariados dentro del crecimiento económico de las naciones en el capitalismo moderno así como las consecuencias negativas para su desarrollo vital a partir de la enajenación del producto de su trabajo.

Este nodo de discusión, configurado a partir de la necesidad de comprensión de la lógicas del naciente modo de producción capitalista y fundamentado en la interpretación del desarrollo de las fuerzas productivas, de la

¹Se entiende por dualización del trabajo como la coexistencia de varias condiciones laborales en un mismo proceso productivo (Neffa 2010).

superespecialización del trabajo y de la creación de una relación social de producción consolidada con base en la propiedad privada y en la posibilidad de la venta libre de la fuerza de trabajo, impulsó el debate más trascendental y candente de la historia reciente de la humanidad: El debate sobre las causas de las desigualdades y la pobreza.

En efecto, como se percibe desde los orígenes de la interpretación teórica moderna, las problemáticas vinculadas al trabajo, han estado fuertemente anidadas a la desigualdad y la pobreza. Este debate que se ha suscrito fundamentalmente a la relación existente entre la condición del asalariamiento, el desarrollo técnico productivo y las condiciones de vida, se ha edificado de manera desigual en virtud de la configuración de los diferentes regímenes de acumulación y modos de regulación que históricamente se han consolidado en el centro y la periferia de producción capitalista (Marini y Sader 1973).

De esta manera, la condición del desarrollo desigual signada por el nivel de incorporación técnica y avance en el desarrollo de las fuerzas productivas, ha tenido fuerte eco en la forma como se desarrollaron las interpretaciones sobre el trabajo en cada territorio. Puede evidenciarse cómo en los países del centro del desarrollo capitalista, particularmente en Inglaterra, Francia y Estado Unidos, el proceso de consolidación de los avances científicos de finales de siglo XIX propios de la segunda revolución industrial², impulsaron la consolidación de varias disciplinas que intentaban interpretar el problema del trabajo y la producción.

Desde el nacimiento de la encuesta moderna (D'Ancona y Ángeles 2012), con el desarrollo de la investigación de los 17 volúmenes de *Labour and Life of the People of London*(1902) impulsados por Macmillan entre 1889 y 1897, pasando por las magistrales interpretaciones de Max Weber en su emblemática tesis doctoral titulada *La ética protestante y el espíritu del capitalismo*(1905) donde se develan las formas de configuración simbólica de la ética laboral moderna a partir del ascetismo protestante, hasta las revolucionarias técnicas de gestión del trabajo desarrolladas por Frederick Taylor en su difundido libro *La Administración Científica* (1911), se evidencia el papel central que jugó el desarrollo de las fuerzas productivas y el trabajo dentro de la configuración y el impulso intelectual en los países desarrollados.

En contraste, la periferia capitalista se observó un exiguo desarrollo intelectual. Los principales debates de principio de siglo XX en América Latina circundaban aún por los linderos del ideologismo político de mediados del siglo XIX. En efecto, los insignificantes debates sobre las condiciones de trabajo y de explotación de los asalariados no tuvieron gran viabilidad, no solamente por el timorato avance del proceso de industrialización y de consolidación del mercado interno, sino por las limitaciones impuestas por una cultura enquistada en el pasado y configurada desde una lógica meramente hacendataria. Fueron pocos lo que se atrevieron hablar de la condición de la naciente clase obrera y de las clases campesinas explotadas por los terratenientes, sin embargo, tal vez con excepción del peruano Mariátegui, sus narrativas no superaban una visión romántica y plenamente adoctrinada del pueblo.

Esta disparidad analítica continuó durante un considerable periodo de tiempo, mientras que en nuestra América buscábamos formas de impulsar los motores productivos con el objetivo de modernizarnos, es decir, de saltarnos los tiempos de la historia a partir de fórmulas y modelos pocas veces efectivos, en los Estados Unidos y la renaciente Europa de la posguerra se desdoblaba el modelo productivo fordista y el conocido Estado de Bienestar, suficientemente robustos para a morigerar los efectos de negativo de la explotación capitalista en el seno de sus naciones.

Durante este lapso de tiempo, comprendido entre la década de los 40's y principio de la década de los 80's del siglo XX, se desarrollaron corrientes teóricas que obedecía al contexto socio histórico y económico de cada nación, sin embargo, se guiaron a partir del derrotero de la sociedad industrial fundamentada en el pleno empleo y la expansión de la relación salarial. En el caso latinoamericano, se evidenció el surgimiento de los discursos desarrollistas y de la modernización guiados por Gino Germani así como la corriente crítica de la Dependencia liderada por Ruy Mauro Marini y Theotonio Dos Santos, la cual analizó la fisura histórica entre el centro y la periferia a partir del aumento paulatino y sistemático de la de la brecha tecnológica, la

²Desarrollo de la petroquímica y la incorporación de la cadena de producción

reprimarización productiva y de la intensificación y superexplotación del trabajo.

Como es evidente, la óptica desarrollada desde América Latina, se apropió de un discurso eminentemente estructuralista que dejó de lado los factores materiales y los actores directos del proceso de producción. En consonancia, desde la escuela estructural funcionalista Norteamérica se privilegió la acción racional y desde la escuela francesa, se realizaron esfuerzos por comprender los procesos del trabajo en virtud de su anclaje al modelo estructural.

Con el advenimiento de la crisis del petróleo de 1973 y el impulso de la teoría neoclásica así como del libre mercado, se genera un proceso de apertura de la economía mundial, en el cual el modo tradicional de producción industria empieza a sufrir un fuerte impacto. Los procesos de deslocalización productiva (maquila y producción en red) conjugados con los cada vez más fuertes proceso de globalización, generaron un fuerte remesón en las formas como se comprendía el mundo productivo y las dinámicas del trabajo.

Enrique de la Garza analizó como el proceso de la reestructuración productiva consolidó un cambio de mentalidad en los estudios laborales en América Latina. A diferencia de los años anteriores, en los cuales el trabajo era comprendido como una consecuencia pasiva del capitalismo o visto sólo desde la óptica de la historia de la movilización política de la clase obrera, con el desarrollo de las de las perspectiva posfordistas (neoinstitucionalismo, neoshupeterianismo, regulacionismo y especialización flexible) y la imposibilidad de estas explicar la forma como mutaba el capitalismo en los territorio periféricos, surge una corriente autóctona en los años 90's que interpreta el proceso de transformación productiva con base en las transformaciones laborales territorializadas y fundamentalmente en el proceso del trabajo.

Sin embargo, el debate de la transformación de los modelos productivo como consecuencia del proceso de reestructuración productiva no es tema que ha venido siendo tratado en la literatura internacional desde hace varios años. Manuel Castells analiza el proceso de flexibilización como un proceso que obedece y se deriva de un cambio radical en la lógica organizativa empresarial a partir de la década de los 80's. Señala que este cambio se materializó por la caducidad del antiguo modelo vertical y monolítico de la industria moderna fordista, el cual le dio paso a un esquema más flexible y adecuado al mercado, el toyotismo. En este sentido, Castells sitúan cinco elemento de dicho cambio de lógica que en ultimas estructurarían el funcionamiento de empresas red -esta entendida como una institución económica que, ostentando medios autónomos, se plasma como herramienta para llegar a un fin determinado de una empresa matriz-: a) una gran fragmentación tanto en la organización de la producción como en los mercados, b) cambios encaminados a una difusión de la tecnología de la información, c) transformación de la adaptación de los cambios generales, tanto en el plano económico-institucional como tecnológico, d) una redefinición de los procesos de contratación y trabajo, y por ultimo e) una operativización del conocimiento y del procesamiento de la información, como mecanismo imprescindibles para las organizaciones que funcionan dentro de una economía informacional global.

Las transformaciones guiadas por el cambio de lógica han generado un cambio de la producción en serie a la producción flexible basada en la subcontratación de empresas red por medio de la revolución de los medios de gestión, es decir, la adecuación de los ejes rectores del toyotismo como una ciencia de la subcontratación. Resaltándose en este sentido el impulso de los vínculos inter empresariales en los cuales se crean interconexión entre una diversidad de empresas.

Para Neffa (2010) la precariedad se encuentra en el corazón de la informalidad por el proceso mismo de flexibilización del trabajo. Plantea la existencia de una distinción entre lo informal e informalidad, ya que la primera comprende el fenómeno de contratación atípica y la segunda el incumplimiento de normas de protección a los trabajadores. Neffa, al retomar a David Harvey en su concepto de la acumulación por desposesión, señala que la precarización de los derechos laborales tiene un función determinada: El aumento de la tasa de ganancia mediante la flexibilización, es decir, la reducción máxima de los costos de mano de obra en contra posición de la maximización de la extracción del plus valor absoluto del trabajador.

Para Castel (1997), quien parafrasea a Sebastien Vauban, el problema de la precariedad - más allá de la consecuente teorización abstracta económica- excede el fenómeno de la baja salarial que determina la miseria

material del trabajado, esta también implica otras particularidades de esta condición como son la inestabilidad del empleo, la búsqueda de ocupaciones provisionales, la intermitencia de los tiempos de trabajo y de no empleo. La vulnerabilidad según Castel pone al pobre a una corta distancia de la indigencia, siendo esta vulnerabilidad y precariedad extendida a la condición del pueblo mientras que la condición general abstracta se plantea como boyante.

El concepto de precarización se somete a debate como un proceso general que no solamente se plantea en el marco de una dualidad de mercado laboral sino como un todo complementario que, en el devenir de auge y crisis del empleo, va tomando diferentes matices y relevancias según sea el caso. “En realidad, un asalariamiento precarizado revitaliza la idea del capitalismo salvaje, de un régimen neoconservador, en la línea sugerida por Bourdieu y Castel, de una sociedad capitalista con un Estado fuerte en el control o la dominación y el uso de la represión, pero mínimo en la regulación de los agentes económicos y en las transacciones de los mercados, incluyendo el mercado laboral” (Urrea 1999).

2.1. Marco muestral

En la medida que la población objetivo hace parte de instituciones económicas privadas, guiadas por la racionalidad instrumental y el cálculo consto beneficio, se asume que dentro de sus estructuras de racionalización de los procesos del trabajo existen registros y bases de datos concretas del personal. En virtud de ello, se puede predecir de manera breve que el marco muestral de la investigación sería de lista, estando anudado a los desarrollos y las precisiones que cada una de las empresas detenten en clave de sus bases de datos de empleados directos y de empresas tercerizadas.

2.2. Tipo de Muestreo

Ya que se presume la existencia de marcos de listas solidos por parte de las empresas, se plantea realizar un muestreo estratificado proporcional que nos permita garantizar “una probabilidad igual de selección para todos los estratos” (D’Ancona y Ángeles 1999), en virtud de su peso diferencial en el número de trabajadores. Ahora pues, vemos la necesidad de complementar este muestreo con la aplicación de Muestreo con Probabilidad Proporcional al Tamaño (PPT), con el objetivo de extraer las riquezas representativas de los estratos existentes.

Como tal, el uso del PPT se vincula regularmente al auscultamiento de posibles relaciones proporcionalmente diferenciales entre las variables auxiliares (en este caso la condición de trabajador directo o indirecto) y la variable de interés del estudio (Condición de los trabajadores en las empresas). En este sentido, asigna probabilidades desiguales y a la vez proporcionales a la variable auxiliar, con el objetivo de reducir la variabilidad entre la variable auxiliar y la variable de interés. La bondad del ajuste que realiza el PTT se encuentra en la posibilidad evidenciar de manera directa cierto tipo de relaciones que quedarían ensombrecidas y vilipendiadas por una explicación poco clara atribuible al azar. El problema que se tiene para diseñar a partir del PTT es que se adolece en el actual momento del marco muestral que permitiría asignar el tamaño de la muestra en función de la variable auxiliar presente en el conjunto de los estratos-empresas.

Sin embargo, se puede plasmar una tentativa parcial de diseño muestral que incorpore los principio del muestreo estratificado proporcional. Para este propósito utilizaremos la fórmula de cálculo de muestra que a continuación se presenta:

$$n = \frac{def f \frac{Z^2 pq}{e^2}}{\frac{1 + def f \frac{Z^2 pq}{Ne^2}}{1 - Pnr}} \quad (1)$$

Donde

n : Tamaño de la muestra.

$def f$: Efecto del diseño

Z : Valor de la distribución Normal estándar para un nivel de confianza del 95 %

p : Proporción de unidades en la población con característica de investigación.

$q : 1 - p$
 e : Error máximo absoluto esperado.
 N : Tamaño poblacional.
 Pnr : Porcentaje de no respuesta.

Por su parte, tomaremos el valor de la varianza de la variable trazadora de la investigación (forma de vinculación del trabajador con la empresa matriz) para calcular de manera más ajustada el posible tamaño de la muestra. La varianza (acierto) de la variable dicotómica trazadora del proyecto calculada con base en la prueba piloto es igual $p = 0.7497$, en oposición a 0.2503 que representa el rechazo producto de $q = 1 - p$.

TABLA 1: Tamaños de muestra por empresa

Empresa	Número de Trabajadores	Proporción Trabajadores/Empresa	Tamaño de muestra
Votorantin	2200	57.89 %	155.74
Diacó-Gerdau	600	15.79 %	42.47
Sidenal	1000	26.32 %	70.79
TOTAL	3800	100.00 %	269

Fuente: Los Autores, 2018

Luego de realizar el cálculo de la muestra con un nivel de confianza del 95 % y un error esperado del 5 %, proseguimos a calcular la proporción de participación de cada una de las empresas dentro de la misma. La proporción esta signada por la razón entre los trabajadores de cada empresa sobre el total de trabajadores del conjunto de las empresas, dicha proporción de participación es multiplicada por el cálculo de la muestra total, dando como resultado el número el tamaño muestral del estrato.

Por su parte, para realizar las pruebas de fiabilidad y consistencia del instrumento, se eligió El Coeficiente Alfa de Cron Bach. Este coeficiente tiene una larga tradición en el campo estadístico y de la investigación de ciencias sociales. El Coeficiente Alfa de Cron Bach produce valores que oscilan entre 0 y 1, su ventaja es que no hay que dividir los ítems en dos mitades, Es un índice de consistencia interna que se utiliza para medir la fiabilidad del cuestionario, puede ser calculado manualmente o en cualquier software estadístico; de dos formas: 1) Mediante la varianza de los ítems o 2) Mediante la matriz de correlación.

El coeficiente Alfa de Cronbach es producto del cociente entre las sumas de varianzas de los diferentes ítems, divididos entre la varianza total, ponderada por el número de ítems de escala. Su expresión matemática es la siguiente:

$$\alpha = \frac{n}{n - 1} \left(1 - \frac{\sum S_i^2}{S_t^2} \right) \quad (2)$$

Donde

n : Número de elementos o ítems de la prueba.
 S_i^2 : Varianza de cada uno de los ítems.
 S_t^2 : Varianza de las puntuaciones totales de la prueba.

El valor mínimo aceptable para el coeficiente alfa de Cronbach es 0.7; por debajo de ese valor la consistencia interna de la escala utilizada es baja (Celina y Campo, 2005). Este valor manifiesta la consistencia interna, es decir, muestra la correlación entre cada una de las preguntas; un valor superior a 0.7 revela una fuerte relación entre las preguntas, un valor inferior revela una débil relación entre ellas. Sin embargo, existen otras interpretaciones que sitúan el límite de fiabilidad en virtud del nivel del nivel de sistematización y uso de las preguntas, situando una buena puntuación para nuevos grupos de preguntas en 0.6 y para preguntas ya calibradas en puntuaciones superiores a 0.8 (Cortina 1993).

3. Metodología

El desarrollo del proyecto se orientó a la construcción y validación de un cuestionario que permitiera realizar la medición de las condiciones socioeconómicas y laborales de los trabajadores fabriles del acero que trabajan en las principales industrias del corredor Industrial del departamento de Boyacá. Para este propósito, se planteó el desarrollo parcial del Modelo Genérico del Proceso de Producción Estadística consolidada por las Naciones Unidas en su versión 5.0, que se fundamenta en procesos iterativos para el desarrollo estadístico que permiten una dinámica continuada de mejoramiento tanto de los productos plasmados desde la metodología como del mismo instrumento de producción.

A pesar que la versión 4.0 de GSBPM desarrollada en el año 2009, hace un mayor énfasis en la construcción, producción y validación de la encuesta dentro del proceso de producción estadística, preferimos la utilización de la última versión ya que fortalece el enfoque basado en procesos por sobre el temático, factor que beneficia y robustece el enfoque tras disciplinar que desarrollamos. Por otra parte, cuando hablamos del uso parcial de la metodología, nos referimos a que utilizamos sólo ciertas fases del proceso referidas al diseño la investigación y la elaboración del instrumento de recolección de los datos.

En este sentido, retomaremos particularmente la fase dos y tres del conjunto de ocho fases que componen el GSBPM 5.0. La fase dos referida al diseño se encuentra compuesta por seis subprocesos, de los cuales ejecutamos los tres iniciales: a) diseño conceptual, b) diseño de la recolección y c) determinación del marco muestral y tipo de muestreo. En relación a la fase tres referida a la elaboración del proceso tomamos tres subprocesos de los siete existentes: a) Construcción de instrumentos de recolección, b) Construcción o mejora de componentes y c) Prueba piloto del proceso estadístico.

3.1. Diseño

3.1.1. Diseño conceptual

Para desarrollar el proceso de medición se construyeron cuatro líneas de análisis gruesas que nos permitirán identificar las condiciones de la población trabajadora que se pretende tipificar. En primer lugar, se edificó una información socio-demográfica básica; en segundo lugar, se construyó una línea orientada a las Condiciones Socio-Económicas y de Seguridad Social de los trabajadores; en tercer lugar, se esbozó la línea gruesa del formulario con base en las Condiciones Laborales de los trabajadores; por último, se planteó una línea complementaria de análisis compuesta por las condiciones del hogar en términos demográficos, materiales y de acceso a servicios públicos fundamentales.

Información Básica

Esta primera línea buscó dar claridad sobre conceptos básicos socio-demográficos de la población trabajadora fabril del acero del corredor industrial de Boyacá. En ese sentido se buscó identificar el sexo, el estado civil, la edad, el país de origen, el municipio de residencia, el lugar de ubicación de la vivienda así como el nivel de escolaridad de los trabajadores. Esta información nos brinda la posibilidad de tipificar las características mínimas individuales de la población objetivo.

Condiciones Socio-Económicas y de Seguridad Social

En esta línea de trabajo se intentó perfilar, de manera general, la condición socio-económica de los trabajadores así como la presencia o ausencia de los elementos mínimos componentes del Sistema de Seguridad Social a nivel Nacional. De esta manera, se buscó precisar los niveles de ingreso individual y gasto colectivo, la posibilidad de existencia de otras fuentes de ingresos diferentes a la condición salarial, la coexistencia de conducta asociativa o cooperativa por parte de los trabajadores y su adscripción o no a los diferentes regímenes de seguridad social existentes (Salud, Pensión y Agencia de Riesgos Profesionales).

La información provista por esta línea de trabajo nos permite identificar, de manera integral, las condi-

ciones económicas relacionadas con a la capacidad de ingreso así como su relación con la presencia o ausencia de condiciones mínimas legales asociadas a derechos fundamentales de los trabajadores.

Condiciones Laborales

Las condiciones laborales son el eje central de la encuesta. Dentro de esta línea se encuentra la variable trazadora del análisis que se sitúa en el tipo de vínculo que el trabajador detenta con la empresa matriz. Dentro de este bloque temático se incluye también una reseña breve de la trayectoria laboral de los trabajadores y, fundamentalmente los perfiles diferenciados de los trabajadores en virtud de su tipo de vinculación. En ese sentido se intenta observar a partir de las preguntas, si existe algún tipo de relación de los trabajadores directos con el sindicato así como la tipificación de la transformación de las condiciones laborales pretéritas de los trabajadores indirectos. También, se intenta discriminar si existen tratos diferenciados entre los dos tipos de trabajadores, en lo referido a desplazamientos de su lugar de residencia al sitio de trabajo así como en la asignación del turno y el pago de las horas y los turnos extras.

Hogar y servicios públicos

Este cuarto grupo de preguntas tiene como objetivo amplificar la tipificación socio-demográfica y socio-económica del trabajador, en clave de las condiciones materiales de habitabilidad que presenta en relación a la cantidad de personas que componen su hogar. Este grupo complementario, parte de una lectura ampliada de la interpretación de la pobreza, se encuentra influenciado por la perspectiva de las capacidades para la medición de la pobreza desarrollada por la COMISIÓN SARKOSY en el año 2010 (realizado por: Stiglitz, Sen y Fitoussi) y el enfoque del Índice de Pobreza Multidimensional de Oxford, que se oponen de manera directa a una medición fundada exclusivamente en la concepción utilitarista del ingreso (González 2011).

3.1.2. diseño de la recolección

Ya que el objetivo del trabajo versó sobre la construcción y validación del cuestionario que nos permitiera identificar la influencia del tipo de vinculación laboral, el proceso de recolección de la información se ancló a la necesidad de prueba del instrumento creado. En esa dirección, y debido a la brevedad del tiempo para la realización de las pruebas, se determinó que los individuos que se perfilarían para el proceso de evaluación no responderían a la producción de tamaño de muestra acorde al tamaño de la población.

El cuestionario consta de sesenta y seis preguntas y está dividido en cuatro secciones, la primera sección consta de ocho preguntas relacionadas con la información básica del trabajador, la segunda sección hace referencia a las condiciones socio-económicas y de seguridad social del trabajador se compone de catorce enunciados, la tercera sección trata de las condiciones laborales y está compuesta de veintisiete interrogantes y la cuarta sección muestra información relacionada con el hogar y los servicios públicos que posee el trabajador, está compuesta por dieciséis preguntas.

TABLA 2: Distribución de las encuestas por empresa

EMPRESA					
ACERÍAS PAZ DEL RIO		DIACO-GERDAU		SIDENAL	
Directos	Indirectos	Directos	Indirectos	Directos	Indirectos
4	5	3	3	3	3

Fuente: Los Autores, 2018

En ese sentido se consensuó un levantamiento de la información a manera de sondeo, donde se representase las tres empresas que hacen parte del actual estudio. Para tal propósito se perfilaron seis visitas informales a las empresas, proyectadas en los cambios de turno de los trabajadores, donde se propuso recoger no menos

de tres encuestas de cada tipo de trabajador por empresa. Luego de realizarse la recolección de los en las seis visitas planteadas, se consiguieron 22 encuestas distribuidas de la manera indicada en la Tabla 2.

3.1.3. determinación del marco muestral y tipo de muestreo

A pesar que los alcances del estudio no nos permiten realizar una aplicación real en tanto que el objetivo central del mismo se erige entorno a la construcción y validación del instrumento de medición, se ve necesario plantear las bases mínimas para el desarrollo de un proyecto de mayor calado. En consecuencia, es necesario plantear la población objetivo, el marco muestral, el tipo de muestreo preferente así como la posible muestra.

Población Objetivo

La población comprende el conjunto de trabajadores y trabajadoras del ramo industrial del acero, que realizan sus labores contractuales en la producción directa del material férreo desde la explotación o manejo de la materia prima hasta el producto final, dentro de las industrias acereras del Departamento de Boyacá.

3.2. Construcción.

3.2.1. Construcción de instrumentos de recolección

La construcción de instrumento se realizó con base en la discriminación tipológica antes resaltada, la cual tuvo como referente la Encuesta Nacional de Hogares desarrollada por el DANE así como la Encuesta Multipropósito impulsada por la misma entidad. El ensamblaje de las preguntas y su codificación partieron de las recomendaciones bibliográficas (D'Ancona y Ángeles 1999), dándosele centralidad a la redacción clara y amable para el encuestador, así como al necesario balance del contenido del lenguaje teórico dentro del mismo.

También se prestó fuerte atención a la construcción de flujos estructurados que nos permitieran darle continuidad a las líneas analítica construidas, y que permitirán generar bloque internos dentro de las mismas líneas para la tipificación de los encuestado. En este sentido, se construyeron preguntas filtro que discriminaran la necesidad o no de la respuesta del encuestado de cierto tipo de preguntas, en total se desarrollaron 13 flujos diferenciados que nos dan la posibilidad de construir ese mismo número de perfiles de la encuesta.

3.2.2. Construcción o mejora de componentes

Luego de la producción analítica y materia del instrumento de medición proseguimos a desarrollar el proceso validación y fiabilidad de la encuesta. Para este propósito se utilizaron tres técnicas diferenciales: dos vinculadas a la validación pragmática y analítica del instrumento y una destinada la consistencia del mismo.

En clave de la validación, en un primer momento, realizamos la consulta de experto en la temática que aborda el instrumento. Pese a que es deseable el desarrollo de un panel de expertos, las posibilidades materiales de consecución de un número suficiente de personas con las aptitudes y la cualificación en la temática no fue sencilla. En ese sentido, se contactaron dos académicos especialista en sociología del trabajo y sociología de la industria, con amplia trayectoria en la temática y con fuertes conocimientos en metodología de la investigación. No obstante, uno de los académicos contactados no pudo brindar el apoyo solicitado producto de su apretada agenda laboral. En ese sentido, sólo se contó con el apoyo del académico Rafael Joya Valderrama, sociólogo, Magister en Sociología Industrial y en Estudios Culturales, y docente de la UPTC con amplia trayectoria en metodología de la investigación, quien estuvo presto a puntuar el cuestionario en cinco diferentes dimensiones y realizar las respectivas recomendaciones.

En relación a la validación pragmática del instrumento, se efectuó, dentro de la prueba piloto del cuestionario, la recolección de las opiniones de los 22 encuestado con base en dos indicadores: Legibilidad³ y

³La claridad en la exposición de los conceptos planteados en las preguntas.

Autosuficiencia⁴. Indicadores en los cuales se el cuestionario obtuvo una alta puntuación de 4.6 sobre 5 y 4.4 sobre 5 respectivamente.

Por otro lado la información obtenida en las encuestas se tabulo en Excel para luego importar el archivo hasta el software estadístico R (R Core Team 2017), donde se obtuvieron los siguientes resultados.

4. Resultados y conclusión

Los resultados del proceso de producción y de validación del cuestionario desarrollado se expresan, en primer lugar, en el proceso de diseño y desarrollo metodológico ya señalado y, en segundo lugar, evaluación y calificación realizada por el experto así como en los resultados obtenidos con la aplicación de Coeficiente de fiabilidad de Alfa de Cronbach que a continuación se presentan.

4.1. Evaluación del par experto

TABLA 3: Evaluación par experto

Tópico	Observaciones	Puntuación
Pertinencia	La temática tratada por el cuestionario tiene suma relevancia para la comprensión de las condiciones socio-económicas y demográficas del Departamento. Cabe destacar que la industria del acero ha sido un motor crucial del desarrollo productivo y de las transformaciones socio-culturales que ha vivido Boyacá en los últimos 70 años. Por ende, el esfuerzo por construir un instrumento de medición que permita observar los cambios y las vicisitudes de lo viejo y nuevos trabajadores, en el marco de las transformación normativa, representa un valioso aporte a la perspectiva disciplinar de la sociología de la industria, económica y del trabajo.	4.7
Suficiencia	A pesar que la producción de este tipo de instrumentos aporte de sobre medida para interpretación de las relaciones industriales y de las dinámicas del mundo del trabajo, adolece varias dimensiones fundamentales para la ininteligibilidad de las dinámicas que configuran el proceso del trabajo. Se escapan factores como la calificación interna de las empresas dada a los trabajadores, las distinciones socio materiales de la clase obrera en el acceso a bienes de consumo privilegiados (el casino), la diferenciación de los implementos de trabajo y sus efectos en el relacionamiento social (dotaciones, diversidad de calidades), olvidándose también una lectura más profunda de la disputa por el control del proceso del trabajo en el piso de la fábrica.	4.2
Coherencia	En términos del desarrollo del constructo el formulario es un instrumento balanceado que incorpora categorías de análisis propias de algunas ramas interpretativas de la sociología del trabajo.	4.4
Relevancia	El instrumento evaluado tiene una relevancia central para la comprensión del desarrollo productivo regional y el desentrañamiento de las nuevas prácticas laborales y empresariales que se plantean en un momento crucial de cambio productivo.	4.5

⁴La posibilidad de desarrollarse como herramienta autodiligenciable.

Sintaxis	La sintaxis del cuestionario ofrece buenas condiciones para su presentación. No se observaron errores ortográficos ni de articulación sintáctica en la construcción de las oraciones que componen los ítems del cuestionario.	4.8
Semántica	A pesar que intenta manejar un lenguaje claro, el instrumenta peca en algunos tecnicismo que pueden llegar a ser difíciles de comprender por parte de la población de estudio. Se recomienda rebajar un poco el uso del lenguaje para que pueda ser más accesible y claro para el entrevistado	4.1
	TOTAL	4.45

Fuente: Los Autores, 2018

Teniendo como referencia la evaluación dada por el par académico $\frac{4.45}{5}$, consideramos que el cuestionario tiene una buena robustez en términos de su validez conceptual y su proyección analítica. Sin embargo, es menester acoplar la sugerencia realizadas por el experto en clave del desarrollo semántico del discurso y la necesidad para evitar generar errores no esperados de interpretación del encuestado.

4.1.1. Fiabilidad del Cuestionario

Luego de aplicar la prueba piloto a los 22 trabajadores ya reseñados, procedimos a realiza la medición de la fiabilidad y consistencia del instrumento de medición. De iniciada, depuramos del análisis 18 variables que no brindaron las condiciones mínimas de variabilidad y que no podían ser contrastadas a partir del coeficiente. Se debe resaltar que, debido a las condiciones del instrumento, el ejercicio de medición de consistencia no se aplicó de manera general al grueso de las preguntas.

El ejercicio constó en la medición de la consistencia de los subproceso de la línea de trabajo central del instrumento. Es decir, de los cuatro posibles perfiles que se extraen de la línea analítica de las condiciones de trabajo: a) Perfil Laboral, b) Vinculación directa y relación sindical, c) Trabajadores indirectos y d) Horarios y turnos de trabajo.

Perfil Laboral

```

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.32 0.62 0.45 0.45 1.7 0.11 11 1.3 0.45

Lower alpha upper 95% confidence boundaries
0.1 0.32 0.54

Reliability if an item is dropped:
raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
COMIENZO 0.45 0.45 0.21 0.45 NA NA 0.45 0.45
VINCULACION 0.21 0.45 NA NA NA NA 0.21 0.45

```

En relación al perfil labora, se puede observar que el coeficiente de alfa de se encuentran por encima de 0.6, dando señales de una buena consistencia interna en lo referido a los años de inicio de la vida laboral y el tipo de vinculación. El Tiempo de vinculación a la empresa matriz fue eliminado por que desmejoraba de manera considerable la suficiencia de los ítems (-0.4) y fue trasladado al grupo de tipificación de los trabajadores indirectos ya que presentaba en ellos cierta variabilidad, a diferencia de los trabajadores directos donde esta variable, valga la redundancia, adolecía de variabilidad.

Sindicatos y vinculación directa

```

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.31      0.52      0.35      0.35 1.1 0.15 2.6 0.84 0.35

lower alpha upper      95% confidence boundaries
0.03 0.31 0.6

Reliability if an item is dropped:
raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
SINDICATO      0.35      0.35      0.12      0.35 NA      NA 0.35 0.35
OPINION.SINDICATO 0.12      0.35      NA      NA NA      NA 0.12 0.35

```

En la medida que el tipo de vinculación a la empresa matriz y la forma de contratación dentro de la población de los trabajadores directos es invariables, la consistencia de su sub-línea tuvo que ser analizada exclusivamente sobre las variables de pertenencia al sindicato y su opinión sobre el mismo. Como se puede observar, el coeficiente de Cronbach no es lo suficientemente elevado para que se pueda hablar de consistencia (0.52), sin embargo, no se puede descartar la posibilidad de un aumento del mismo ya que es probable que ante una muestra de mayor tamaño cambie la situación de la invariabilidad de las dos primeras variables.

Trabajadores Indirectos

```

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
0.35      0.66      1      0.22 2 0.12 37 27 0.25

lower alpha upper      95% confidence boundaries
0.12 0.35 0.59

Reliability if an item is dropped:
raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
TIEMPO.VINCULACION 0.23      0.62      1      0.21 1.64 0.100 0.202 0.281
TRABAJOS.ANTERIORES 0.13      0.66      1      0.24 1.94 0.082 0.198 0.252
CONTRATO.INDIRECTOS 0.36      0.68      1      0.26 2.14 0.122 0.149 0.281
DURACION.CONTRATO 0.29      0.47      1      0.13 0.88 0.123 0.178 0.097
RENOVACIONES 0.36      0.71      1      0.29 2.42 0.123 0.143 0.318
OTRAS.EMPRESAS 0.36      0.60      1      0.20 1.48 0.123 0.131 0.281
DIAS.PROMEDIO 0.35      0.60      1      0.20 1.48 0.125 0.086 0.235

```

La sub-línea de los trabajadores indirectos es la que, a nuestro parecer, detenta mayor robustez. Dentro de esta se tuvo en cuenta el tiempo de vinculación con la empresa matriz, el tipo de trabajo en trabajos anteriores, la vinculación indirecta, la duración del contrato, el número de renovaciones contractuales, el haber trabajado en otras empresas tercerizadas asociadas a la casa matriz y los días de trabajo promedio. Como se evidencia el coeficiente indica una relativa consistencia (0.66) que preferimos no aumenta-retirando el número de renovaciones- en tanto que dicha variable es central para comprender la movilidad, la poca seguridad jurídica y la incertidumbre que viven los trabajadores tercerizados.

Horarios y turnos

raw_alpha	std.alpha	G6(smc)	average_r	S/N	ase	mean	sd	median_r	
0.49	0.68	0.61	0.41	2.1	0.14	2.2	0.62	0.39	
lower alpha upper		95% confidence boundaries							
0.21	0.49	0.77							
Reliability if an item is dropped:									
	raw_alpha	std.alpha	G6(smc)	average_r	S/N	alpha	se	var.r	med.r
HORARIO.INICIO	0.71	0.72	0.56	0.56	2.5	0.12	NA	0.56	
HORAS.EXTRA	0.32	0.44	0.29	0.29	0.8	0.19	NA	0.29	
TURNOS.EXTRA	0.38	0.56	0.39	0.39	1.3	0.15	NA	0.39	

Por último observamos la sub-línea de horarios y turnos de trabajo. Esta fue la línea que presento mayor complicación en tanto que tuvieron que ser retiradas 4 variables para poder lograr consolidar un fiabilidad considerable (0.68).

5. Conclusiones

Luego de haber realizado el proceso de validación con el par experto y de efectuar el proceso de medición de la robustez y fiabilidad de la parte central del instrumento, es evidente que el instrumento aún no se encuentra acaba. No obstante, presenta buenas características para su desarrollo y su mejoramiento a partir del proceso iterativo de producción y consolidación estadística. Siendo el primer intento de validación de la encuesta y de la medición de su consistencia interna, podemos aducir que obtuvimos muy buenos resultado, satisfaciendo las necesidades de constructo y alcanzado de manera promedio el umbral mínimo de fiabilidad para nuevos instrumentos.

Referencias Bibliográficas

- Bouffartigue, P. y Busso, M. (2010), Precariedad, informalidad: una perspectiva “norte-sur” para pensar las dinámicas del mundo del trabajo, *in* ‘VI ème congrès de l’Association Latino Américaine de Sociologie du Travail’.
- Castel, R. (1997), ‘La metamorfosis de la cuestión social. Una crónica del salariado’, *Editorial Paidós*.
- Cortina, J. M. (1993), ‘What is coefficient alpha? an examination of theory and applications.’, *Journal of applied psychology* **78**(1), 98.
- D’Ancona, C. y Ángeles, M. (1999), *Metodología cuantitativa: estrategias y técnicas de investigación social*, Síntesis.
- D’Ancona, C. y Ángeles, M. (2012), *Fundamentos y aplicaciones en metodología cuantitativa*, Editorial Síntesis.
- González, J. I. (2011), ‘Utilitarianism and measurements of poverty’, *Revista de Economía Institucional* **13**(25), 89–103.
- Marini, R. M. y Sader, E. (1973), *Dialéctica de la dependencia*, Era México.
- Marx, C. (2008), ‘El capital i: Crítica de la economía política (4ta reimpression)’, *México DF: Fondo de Cultura Económica* **1867**.
- Neffa, J. C. (2010), ‘La transición desde los “verdaderos empleos” al trabajo precario’, *E. De la Garza Toledo, & JC Neffa, Trabajo, identidad y acción colectiva* pp. 43–80.

R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

*<https://www.R-project.org/>

Urrea, F. (1999), 'Globalización, subcontratación y desregulación laboral', *Fernando Urrea et al., Innovación y cultura de las organizaciones en tres regiones de Colombia, Bogotá, Colciencias/Corporación Calidad* pp. 219–281.



ANÁLISIS DE SATISFACCIÓN Y APROBACIÓN EN FORMACIONES TÉCNICAS DE UNA EMPRESA SIDERURGICA DE BOYACÁ EN 2017

Especialización en Estadística

LAURA MILENA VARGAS PÉREZ^{1,a}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

Este artículo analiza la satisfacción y la aprobación de los colaboradores por las formaciones recibidas en una empresa siderúrgica de Boyacá durante 2017. Abarca los niveles 1 y 2 del método Return On Investment (ROI) de Jack Phillips (1996). Dado que las bases de datos están compuestas por datos categóricos, se implementa para la técnica estadística de Análisis de Correspondencias Múltiples (ACM), la cual permite categorizar a los colaboradores que presentaron satisfacción e insatisfacción, de la misma manera con los colaboradores que reprobaron (los que aprobaron no son foco de este análisis) o estuvieron ausentes a las formaciones. Esto con el fin de generar planes de acción que permitan a la empresa mejorar en el plan de formación.

Palabras clave: Análisis de Correspondencias Múltiples, Método Return On Investment.

Abstract

This article analyzes the satisfaction and approval of the collaborators for the formations received in a steel company of Boyacá during 2017. It covers levels 1 and 2 of the method Return On Investment (ROI) of Jack Phillips (1996). Since the databases are composed of categorical data, it is implemented for the statistical technique of Multiple Correspondence Analysis (MCA), which allows to categorize the collaborators who presented satisfaction and dissatisfaction, in the same way with the collaborators who failed (those who approved are not focus on this analysis) or were absent from the formations. All this in order to generate action plans that allow the company to improve the training plan.

Key words: Multiple Correspondence Analysis, Return On Investment method..

1. Introducción

La casa matriz de la empresa siderúrgica de estudio, ubicada en el extranjero, adelanta procesos de formación en la empresa mediante 5 escuelas de conocimiento: Gestión, SSMA (Seguridad, Salud y Medio Ambiente), Comportamental, Liderazgo y Técnica. En este trabajo se abordará únicamente la escuela Técnica, la cual se encarga de los conocimientos específicos de cada uno de los procesos tanto teóricos como prácticos, la divulgación de conocimientos se realiza a través de formaciones (capacitaciones) diseñadas y adaptadas a cada una de las temáticas. La metodología a utilizar para realizar la medición de impacto en la formación es el Método Return On Investment desarrollado por Jack Phillips que se divide en 5 niveles de medición de

^aEstudiante de Especialización en Estadística. E-mail: lauramilena.vargas@uptc.edu.co

los cuales en este estudio se abordarán los dos primeros niveles y la empresa sumirá la ejecución de los tres restantes. La razón por la que no se puede desarrollar la metodología completa es porque las bases de datos necesarias para ejecutar los niveles 3, 4 y 5 están en proceso de consolidación.

Mediante este análisis se busca resolver los siguientes interrogantes:

- ¿Cuál es la categorización de los colaboradores con mayor grado de reprobación?
- ¿Cuál es o son los ítems más influyentes para cada componente de la satisfacción?

1.1. Justificación

La necesidad de realizar la medición de formación en la siderurgia, surge del requerimiento emitido por la casa matriz de la empresa, la cual, está interesada en conocer cuál es el retorno sobre la inversión de los programas de formación. Este Retorno se calcula bajo el método Return On Investment de Jack Phillips el cual consta de 5 niveles, el método actúa como un algoritmo en el que hay pasos ordenados que se deben seguir y respetar.

En el levantamiento del estado del arte se encontraron 2 empresas en Colombia que han realizado estudios de medición de impacto en formación hasta el nivel financiero, esto se debe a la exigencia intelectual y de tiempo que implica hacer la investigación, lo que hace que el desarrollo sea un reto que la empresa debe asumir.

El análisis de los niveles 1 y 2 del método ROI arrojará resultados que serán fundamentales para el equipo de planeación y desarrollo en la toma de decisiones del plan de formación del año 2019 e incluso, si es necesario, hacer reformas en el plan de formación del año actual.

1.2. Antecedentes

Los primeros estudios referenciados sobre medición de impacto de capacitación se han realizado desde 1959 por Donald Kirkpatrick quien es uno de los iniciadores en este tema, con el modelo de medición conocido como el modelo pionero de evaluación. Desde su surgimiento ha sido el más utilizado por sus características de simple, flexible y completo. Consta de cuatro niveles, dispuestos en una jerarquía de tal modo que, la evaluación comienza siempre por el nivel I y termina en el nivel IV y cada nivel sirve de base para construir los siguientes. (Qian, Li y Zhou 2008) A partir de este modelo se originan estudios, reformas y avances por parte de otros investigadores quienes generan otros modelos de medición.

En este estudio se aborda el método ROI de Jack Phillips en el que empezó a trabajar cerca de 1970 e hizo público en 1996, quien ha ideado una metodología sencilla para calcular la relación costo-beneficio de un programa de adiestramiento. Su método se basa en el modelo de cuatro niveles de evaluación planteado por Donald Kirkpatrick, pero agrega un quinto nivel de tipo financiero y por medio del cual se calcula el ROI. (Franceschi Ramírez y Hernández Pulido 2007)

En Estados Unidos, la mayoría de las empresas miden el nivel 1 (74%), cerca de la tercera parte el nivel 2 (31%), algunas el nivel 3 (14%) y pocas el nivel 4 (8%) como lo muestra ASTD y Saba 2004. Al respecto Boverie, Sánchez & Zondlo, 1994 comentan: Esto refleja la dificultad de medir el impacto en resultados que aún tiene la capacitación, básicamente por la falta de métodos apropiados para aislar el efecto de la capacitación de la multitud de otras variables que impactan los resultados. (Torres Ordóñez 2005)

En Colombia han sido pocos los trabajos realizados (publicados) bajo la metodología ROI. En Ecopetrol este estudio fue de gran importancia porque contribuyó al cambio de paradigmas, suministrando información relacionada con el costo/beneficio de las inversiones realizadas en el área de recursos humanos. Con este modelo se midieron beneficios tangibles como la disminución de los costos y el aumento de la rentabilidad derivada de los programas de recursos humanos, e intangibles como el incremento de la satisfacción y motivación de los empleados, mejoramiento continuo de la gestión, disminución de los conflictos organizacionales etc., que redundan sobre el mejor desempeño de la organización (López Dukmak 2004)

La empresa TGT Gamma S.A.S. también implementó el método ROI en 2014 (Supelano 2014), donde encontró el valor monetario que su programa de capacitación genera a la empresa, además identificó puntos débiles del programa que sirven de base para generar un plan de acción

2. Referente conceptual

En las empresas, los Recursos Humanos constituyen una fuente de vital importancia, es sabido y obvio que cualquier tipo de actividad está directa o indirectamente generada, organizada, desarrollada y controlada por personas. Por lo tanto, se parte de la base de que sin una estructura humana no existe organización y en consecuencia los recursos humanos son parte esencial en el apartado de gestión (García Parejo y Sánchez 2007). Uno de los principales objetivos del área de Recursos Humanos es la formulación de estrategias organizacionales para cumplir las cuatro funciones básicas: obtener personas, prepararlas, estimularlas y conservarlas (Flores Villalpando et al. 2016)

Se entiende por gestión de Recursos Humanos las actividades encaminadas a obtener y coordinar los R.H. de una organización. Los R.H. representan una de las mayores inversiones de una organización. (En Estados Unidos, un 73 % de la renta nacional se va en remuneración de los empleados). Además de sueldos y salarios, las organizaciones realizan inversiones importantes en Reclutamiento, Contratación y Formación y todo esto debe gestionarse (De la Fuente García, Fernández, Fernández y García 2006). La formación profesional es el conjunto de acciones que permiten a una persona alcanzar y desarrollar los conocimientos indispensables para ocupar un puesto de trabajo y acrecentar las destrezas necesarias para su progreso laboral, con satisfacción de sus necesidades técnicas y humanas y las de la empresa a la cual sirve o servirá (Baquero Corrales 2005).

Actualmente la preocupación por realizar mediciones de impacto en las actividades de formación del personal propuestas por el área de Recursos Humanos está tomando gran importancia en las empresas a nivel general, como lo demuestra la Oficina Internacional del Trabajo (OIT) en su guía de evaluación de impacto: Es necesario conocer el grado en que los programas de formación contribuyen a alcanzar objetivos relativos al empleo y la mejora de las condiciones sociales y económicas de las personas, las empresas y la sociedad en su conjunto.

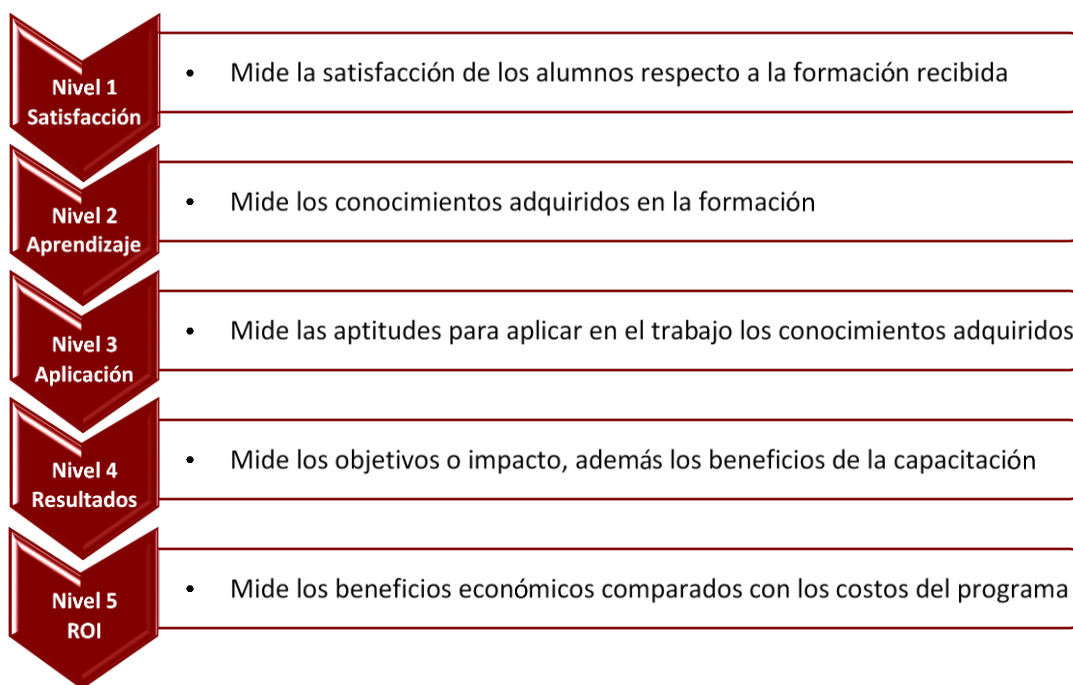


FIGURA 1: Fuente Supelano 2014

Las intervenciones para la formación y el desarrollo de competencias suelen incluir evaluaciones de la satisfacción y el aprendizaje alcanzado. Sin embargo, los resultados y efectos directos, se evalúan con menor frecuencia, aun cuando los responsables de la implementación, las empresas y los propios participantes, están cada vez más interesados en conocerlos (Billorou, Pacheco y Vargas 2011).

La empresa siderúrgica en mención va a la vanguardia de la actualidad, estableciendo como directriz de la organización, la medición de impacto e capacitación para el año 2017. Desde esta perspectiva existe un método que se adecúa a la necesidad de la empresa sobre la que se realiza este trabajo y con el cuál se iniciará el estudio. El método Return On Investment de medición de capacitación (de ahora en adelante método ROI) diseñado por Jack Phillips y el cual toma como base la metodología de medición de Donald Kirkpatrick (cuatro niveles de medición) y adiciona un quinto nivel en el que realiza el análisis financiero.

Basado en la metodología ROI en el presente trabajo se analizarán las bases de datos de satisfacción y aprendizaje del área de formación los cuales corresponden al nivel 1 y 2 de la metodología. *Ver Anexo 1 5*

2.1. Tratamiento estadístico de respuestas basadas en escalas likert

Las escalas Likert discretizan las variables a las que se refieren en un número reducido de valores potenciales. El análisis estadístico de respuestas de un cuestionario basado en una escala Likert consistiría, en principio, en un estudio de datos categóricos. En ocasiones, y con el objeto de ampliar el conjunto de los posibles métodos que puedan aplicarse para tal análisis, se codifican tales datos categóricos mediante números enteros, y se aplican algunos métodos estadísticos para el tratamiento de datos numéricos. Unos y otros resultan bastante limitados, puesto que el número de valores distintos es muy pequeño (de la Rosa de Saa 2012)

Cuando se trabaja con variables categóricas, los datos suelen organizarse en tablas de doble entrada en las que cada entrada representa un criterio de clasificación (una variable categórica). Como resultado de esta clasificación, las frecuencias (el número o porcentaje de casos) aparecen organizadas en casillas que contienen información sobre la relación existente entre ambos criterios. A estas tablas de frecuencias se les llama tablas de contingencia (Pardo Merino y Ruiz Díaz 2005)

2.2. Análisis de Correspondencias Múltiples

El análisis de correspondencias se desarrolla mediante el trabajo sobre dos tablas de datos: la primera tabla contiene las frecuencias respecto a las modalidades (categorías) de dos variables; usualmente se denomina análisis de correspondencias binarias. El segundo tipo de tabla contiene la información sobre diversas variables; el análisis se conoce como de correspondencias múltiples (Díaz Monroy y Morales Rivera 2009)

Las filas de estas tablas son los objetos o individuos y las columnas las modalidades de variables categóricas. Es el caso de las encuestas, donde las filas son individuos, grupos humanos o instituciones; y las columnas, modalidades de respuesta a las preguntas formuladas en el cuestionario o instrumento. El análisis de correspondencias múltiples es un análisis de correspondencias simple aplicado no solo a una tabla de contingencia, sino a una tabla disyuntiva completa (Díaz Monroy y Morales Rivera 2009)

Generalidades

- El ACM trata de analizar, describir y representar gráficamente la información contenida en una tabla de distribución conjunta de datos dispuestos en filas y columnas: sus correspondencias (asociaciones)
- Es una técnica destinada al análisis de la relación de variables cualitativas, tratadas como nominales
- En la representación gráfica cada categoría o valor de la variable se representa como un punto en el espacio: puntos-fila y puntos-columna
- Las proximidades geométricas entre puntos-fila y puntos-columna traducen las asociaciones estadísticas entre filas y columnas (López Roldán y Fachelli 2015)

Tablas de datos

El análisis de correspondencia múltiple tiene como base de estudio la matriz de datos X con n individuos ($i=1,2,\dots,n$) y q variables ($k=1,2,\dots,q$). Esta matriz o tabla de datos no es tratable vía análisis de correspondencias múltiples; pues la suma de estos números en filas o en columnas no tienen ningún sentido. Una salida para el análisis de esta tabla es una recodificación.

Tabla Disyuntiva

Esta recodificación se logra cruzando los individuos con las combinaciones de modalidades para cada una de las preguntas. Mediante el uso de variables indicadoras se convierte una tabla múltiple en una tabla de doble entrada. Suponga, en general, una tabla con k -variables (o preguntas) donde cada una de las cuales tiene p_k modalidades o categorías, $p_k = \text{número de categorías de la variable } k$. Se asocia una variable indicadora por modalidad dentro de cada variable o entrada de la tabla. La codificación dada por p hace corresponder tantas variables binarias como modalidades tenga la variable categórica. El total de modalidades es igual a

$$p = \sum_{k=1}^q p_k \quad (1)$$

Para un individuo particular se codifica con uno (1) si el individuo posee el atributo de la respectiva modalidad y con cero (0) en las demás modalidades de la misma variable, pues se asume que las modalidades son excluyentes. Resulta entonces una matriz X de tamaño $(n \times p)$ formada por bloques columna, cada uno de los cuales hace referencia a una variable registrada sobre los n individuos. (Díaz Monroy y Morales Rivera 2009)

La suma en cada una de las filas es constante e igual a k , mientras que la suma en las columnas n_j suministra el número de individuos que participan en cada una de las q modalidades. La tabla o matriz Z con n -filas y p -columnas describe las p -respuestas para los n -individuos a través de un código binario (0 o 1) y se le llama tabla disyuntiva completa. Esta tabla es la unión de k tablas (una por pregunta).

$$Z_{ij} = \begin{cases} 1 & \text{si el } i\text{ésimo individuo esta en la categoría } m \text{ de la variable } k, \\ 0 & \text{si el } i\text{ésimo individuo no esta en la categoría } m \text{ de la variable } k. \end{cases} \quad (2)$$

Ver Anexo 2 5

Tabla de Burt

Para cada pregunta o variable, sus p_k respuestas o modalidades permiten particionar la muestra en máximo p_k clases. Por ejemplo, para dos preguntas se pueden hacer dos particiones del conjunto de individuos, con lo cual se obtiene una tabla de contingencia. El análisis se puede generalizar a j particiones, donde $j \geq 2$.

A partir de la tabla disyuntiva completa Z , se construye una tabla simétrica B de tamaño $(p \times p)$ que contiene las frecuencias para los cruces entre todas las j variables. Esta tabla es $B = Z'Z$ la cual se le conoce como tabla de contingencia Burt asociada a la tabla disyuntiva completa Z (Díaz Monroy y Morales Rivera 2009)

Sobre la diagonal de la tabla de Burt B , se han insinuado matrices diagonales. Estas se notan por $D_k = PZ_k'PZ_k$; $k = 1, \dots, q$ y son matrices de tamaño $(p_k \times p_k)$. Dichas matrices son diagonales puesto que un individuo no puede tener simultáneamente dos o más modalidades en una misma pregunta o variable. Los términos de la diagonal son las frecuencias de las modalidades de la pregunta k ; es decir, es el número de individuos por modalidad en la pregunta j . Las matrices fuera de la diagonal de B son las tablas de contingencia entre las respectivas variables. Se nota por D a la matriz diagonal de tamaño $(p \times p)$; es decir, sobre la diagonal están las frecuencias correspondientes a cada una de las modalidades. (Díaz Monroy y Morales Rivera 2009)

Conceptos Generales

- **Vectores propios:** son los factores, se extraen un total de $\min\{I, J-1\}$

- **Valores propios:** expresan la inercia relativa (la varianza explicada) de cada eje
- Criterios del número de factores a retener
 - ◊ Considerar el número de ejes que acumulan en torno al 70 % de la inercia total
 - ◊ Representar gráficamente los factores y valores
 - ◊ Interpretabilidad y pertinencia conceptual de los ejes obtenidos
- La **contribución absoluta** de cada punto a la inercia explicada por el eje factorial
- La **contribución relativa**, la correlación entre puntos-fila y ejes factoriales, mide la contribución relativa del factor o eje en la posición de una modalidad, la calidad de su representación (López Roldán y Fachelli 2015)

Ver Anexo 3 5

A partir de la tabla disyuntiva completa se procede a realizar el análisis de correspondencias para ubicar la nube de puntos junto a los factores y las distancias de asociación Ji-cuadrado

Tabla 1. Nubes de puntos - perfiles

	Coordenadas	Peso	Métrica	Distancia χ^2
PERFILES FILA (Individuos)	$\frac{Z_{ij}}{Z_{i\bullet}} = \frac{Z_{ij}}{q}$	$\frac{Z_{i\bullet}}{nq} = \frac{q}{nq} = \frac{1}{n}$	$diag\left(\frac{nq}{Z_{i\bullet}}\right)$	$d^2(i, i') = \frac{n}{q} \sum_{j=1}^p \frac{1}{Z_{i\bullet}} (Z_{ij} - Z_{i'j})^2$ $= \frac{n}{q} \sum_{j \in M_{i'}} \frac{1}{Z_{i\bullet}}$
PERFILES COLUMNA (Categorías)	$\frac{Z_{ij}}{Z_{\bullet j}}$	$\frac{Z_{\bullet j}}{nq}$	$diag\left(\frac{nq}{Z_{i\bullet}}\right)$ $= diag\left(\frac{nq}{q}\right)$ $= diag(n)$	$d^2(j, j') = n \sum_{i=1}^n \frac{1}{Z_{i\bullet}} \left(\frac{Z_{ij}}{Z_{\bullet j}} - \frac{Z_{i'j}}{Z_{\bullet j'}} \right)^2$

(Universidad de Costa Rica)

Inercia

Centro de gravedad de la nube de modalidades:

$$G = \frac{1}{n}, \dots, \frac{1}{n}$$

$$d^2(j, G) = \frac{n}{Z_{\bullet j}} - 1$$

Inercia de la categoría j

$$I(j) = \frac{Z_{\bullet j}}{nq} d^2(j, G) = \frac{Z_{\bullet j}}{nq} \left(\frac{n}{Z_{\bullet j}} - 1 \right) = \frac{1}{q} \left(1 - \frac{Z_{\bullet j}}{n} \right)$$

Inercia de la variable k

$$I(k) = \sum_{j=1}^{p_k} I(j) = \sum_{j=1}^{p_k} \frac{1}{q} \left(1 - \frac{Z_{\bullet j}}{n} \right) = \frac{1}{q} (p_k - 1)$$

Inercia total (Suma de los valores propios)

$$I = \sum_k I(k) = \sum_k \frac{1}{q} (p_k - 1) = \frac{1}{q} (p - q) = \frac{p}{q} - 1$$

(Universidad de Costa Rica)

Interpretación Gráfica

- Buscar las categorías con mayor contribución absoluta
- De estos se distinguen entre los positivos y los negativos para definir las polaridades del eje
- Se estudia la calidad de la representación de los puntos, los valores más altos de contribución relativa
- Interrelacionan los ejes para dar cuenta de la estructura de relaciones teniendo en cuenta el orden jerárquico de cada eje
- Una categoría que coincide con el perfil medio se ubicará en el centro del espacio cercano al origen (“tipo ideal promedio”). Si se aleja difiere de este promedio.
- Si dos filas (o columnas) tienen perfiles similares se situarán próximos en el espacio.
- Equivalencia distribucional: las distancias entre dos modalidades no se alteran si se juntan. Criterio de recodificación. (López Roldán y Fachelli 2015)

Ver Anexo 4 5

En resumen, el análisis de correspondencias múltiples se dirige a buscar aquellas variables o factores “cercanas” (altamente correlacionadas) a todos los grupos de modalidades. El factor *F* representa el primer factor común al conjunto de variables categóricas iniciales. Los demás factores se obtienen con la condición de ortogonalidad sobre los anteriores. Los factores *F1, F2, ..., Fk*, ubicados en el espacio de las modalidades, son los ejes en el espacio de los individuos; de tal forma que su proyección sobre estos ?nuevos? ejes retienen la máxima variabilidad. (Díaz Monroy y Morales Rivera 2009)

3. Diseño Metodológico

El enfoque del proyecto es cuantitativo en el cual se realizará una investigación descriptiva, sobre las bases de datos de Satisfacción y Aprendizaje del área de capacitación de una empresa siderúrgica boyacense.

Las bases de datos contienen la información recolectada durante 2017 de los trabajadores de la siderurgia en mención, quienes respondieron dos cuestionarios, uno de satisfacción y otro de aprendizaje cada vez que asistieron a formaciones, sin embargo, teniendo en cuenta la delimitación del alcance de este trabajo,

la población objetivo serán únicamente los trabajadores que asistieron a formaciones de la escuela técnica durante el año 2017.

La información consignada en las bases de datos se obtuvo por requerimiento del área de recursos humanos a los formadores de cada uno de los cursos, quienes una vez terminadas las clases debía acercarse a la oficina de R.H. para entregar:

- Las listas de asistencia debidamente diligenciadas y firmadas por los colaboradores
- Las evaluaciones de aprendizaje de cada colaborador calificada
- Las encuestas de satisfacción debidamente diligenciadas, en este caso particular las encuestas manejan anonimato, por lo que la verificación se hacía verificando que la cantidad de asistentes a la clase fuese la misma cantidad de encuestas de satisfacción entregadas.

Las bases de datos fueron tabuladas por practicantes universitarios que trabajaron en el área durante el año, quienes se encargaban de actualizar y controlar el acceso a bases de datos garantizando el óptimo manejo de las mismas. Durante el procesamiento y análisis estadístico se parte de que éste debe estar alineado a los datos categóricos dado que las dos bases manejan este tipo de datos. Se generan la respectiva tabla de contingencia las cuales son analizadas para identificar frecuencias altas en las categorías, de esta manera se van identificando variables que arrojan información relevante en el estudio. Seguido de este primer análisis exploratorio se procede a implementar la técnica estadística “análisis de correspondencias múltiples” el cuál arroja las categorías de variables más asociadas a las variables interés de estudio, es decir, revela las categorías que más se asocian a los colaboradores que aprobaron o reprobaron las formaciones, también, las categorías que más se asocian a los colaboradores que estuvieron satisfechos o insatisfechos.

3.1. Variables

3.1.1. Base de datos de satisfacción

- **MES:** Indica el mes en que el colaborador asistió a la formación
- **FORMACIÓN:** Sigla de cada formación a la que asistió el colaborador acerca de la cual respondió la encuesta de satisfacción.
- **COMPONENTE FORMACIÓN:** Los ítems 1, 2, 3 y 4 califican si:
 - ◊ El contenido de la formación fue el adecuado.
 - ◊ Se desarrollaron conceptos claros durante la formación gracias a la estrategia metodológica.
 - ◊ La formación promueve estrategias de innovación y aplicación del aprendizaje a la práctica.
 - ◊ Se utilizaron materiales y recursos adecuados.
- **COMPONENTE INSTRUCTOR:** Los ítems 5, 6, 7 y 8 califican si el instructor:
 - ◊ Mostró dominio del tema y manejo del grupo de trabajo
 - ◊ Motivó la participación de los asistentes
 - ◊ Ayudó a la comprensión de los temas con ejemplos claros, remarcó lo importante y realizó conclusiones
 - ◊ Utilizó un lenguaje, ritmo de exposición y tono de voz apropiado.
- **COMPONENTE LOGÍSTICA:** Los ítems 9, 10 y 11 califican si:
 - ◊ La información previa sobre el desarrollo de la formación fue adecuada y oportuna.
 - ◊ Se respetaron y cumplieron las fechas, tiempos y horarios establecidos
 - ◊ El tiempo programado para la formación fue adecuado

3.1.2. Base de datos de aprendizaje

- **REGISTRO:** Código de identificación de cada colaborador
- **DIVISIÓN:** División de la empresa a la que pertenece el colaborador
- **GÉNERO:** Describe el género del colaborador que asiste a cada formación:
- **EDAD:** Rango de edad en el que se encuentra ubicado el colaborador
- **NIVEL ACÁDEMICO:** Indica el nivel de estudios de cada colaborador
- **JORNADA:** Indica la jornada en la que el colaborador estuvo en clase
- **DÍA SEMANA:** Especifica el día de la semana en que el colaborador asiste a la formación
- **MES:** Indica el mes en que el colaborador asistió a la formación
- **FORMACION:** Sigla de cada formación a la que asistió el colaborador acerca de la cual respondió la encuesta de satisfacción
- **INSTRUCTOR:** Indica el ID asociado a cada formador mediante el cual se identifican los mismos.
- **CLASE:** Modalidad de la clase a la que se presenta el colaborador
- **ASISTIÓ:** Indica asistencia a las sesiones a las que fue programado el colaborador
- **APROBÓ:** Indica aprobación o reprobación de la formación a la que asistió el colaborador
- **INTENSIDAD HORARIA DEL CURSO:** Duración de la formación en horas

4. Resultados

A manera global se hace análisis del estado actual de satisfacción y aprendizaje:

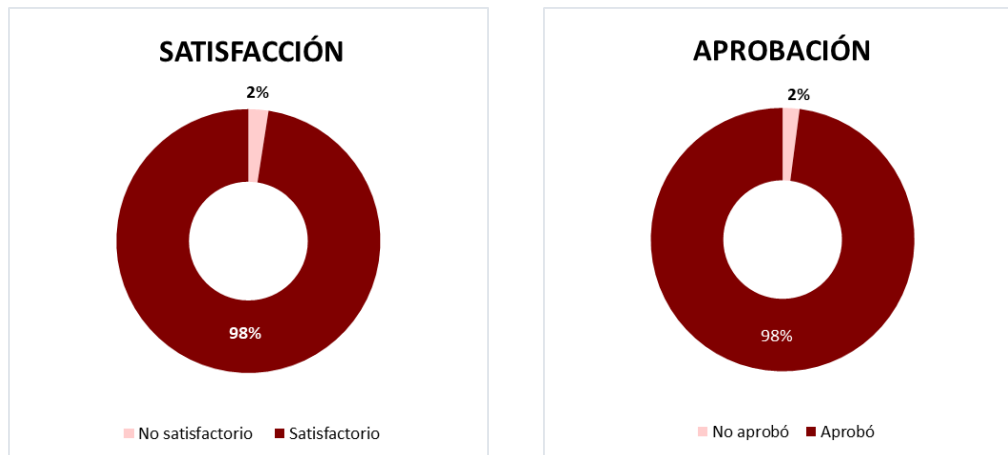


FIGURA 2: Gráfica general. Fuente El autor 2018

Se evidencia un índice de Satisfacción alto a nivel empresa, sin embargo, es importante identificar los puntos débiles causantes de insatisfacción. De manera similar, el índice de reprobación es bajo.

Dado que el fin último del análisis en los niveles 1 y 2 del método ROI es identificar falencias para generar planes de acción que permitan mejorar el desempeño del área de capacitación, se procede a realizar el análisis por medio de Análisis de Correspondencias Múltiples para categorizar a los colaboradores que presentaron insatisfacción y reprobación.

Dado que las variables que se presentan en un análisis de correspondencias múltiples deben ser del mismo “tipo” se agrupan las variables como sigue

4.1. SATISFACCIÓN

4.1.1. ACM Satisfacción Vs Formación

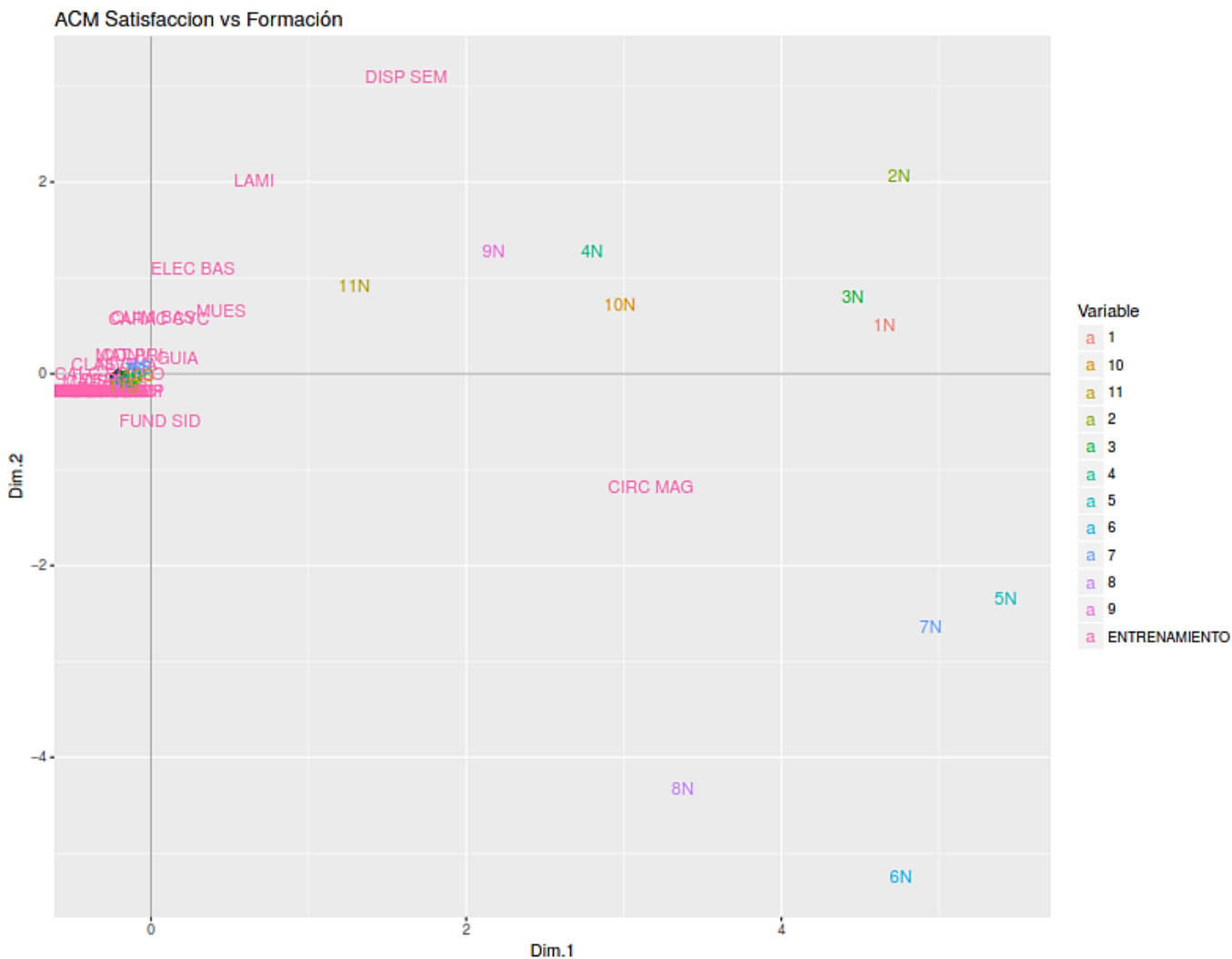


FIGURA 3: Satisfacción Vs Formación. Fuente El autor 2018

Gráficamente se observa que, en términos generales, la satisfacción del personal es alta pues se evidencia la concentración de formaciones en torno a los ítems que resultaron satisfechos, sin embargo, las formaciones que tienen mayor asociación a insatisfacción son Circuitos y magnetismo, Dispositivos semiconductores, Laminado, Electricidad básica.

Tabla 3. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
3N	1,14E+07	9,82E+05	6,48E+04	4,83E+06	4,31E+06
2N	1,11E+07	5,50E+06	5,66E+05	1,80E+06	2,98E+06
1N	1,07E+07	3,33E+05	5,43E+06	1,62E+06	9,24E+06
7N	1,01E+07	7,47E+06	1,46E+05	7,00E+05	2,05E+06
5N	9,67E+06	4,71E+06	1,73E+04	4,76E+06	4,85E+06
CIRC MAG	6,62E+06	2,39E+06	1,27E+04	3,83E+05	5,92E+05
⋮	⋮	⋮	⋮	⋮	⋮
11S	5.333e-01	6.952e-01	4.860e-01	2.425e+00	2.771e-01

Como se observa en la tabla, las categorías que mayor contribución tienen a la varianza son los ítems que resultaron no satisfechos, entre las formaciones se destaca Circuitos y magnetismo, la cual se observa claramente está asociada con la insatisfacción, también Dispositivos semiconductores, matemática básica, tornillería, laminado, electricidad básica.

Tabla 4. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.413409792	12,0997988	12,0998
dim 2	0.158020505	4,6249904	16,72479
dim 3	0.122985106	3,5995641	20,32435
dim 4	0.116451586	3,4083391	23,73269
dim 5	0.102382827	2,9965706	26,72926
dim 6	0.089874631	2,630477	29,35974
⋮	⋮	⋮	⋮
dim 41	⋮	⋮	⋮

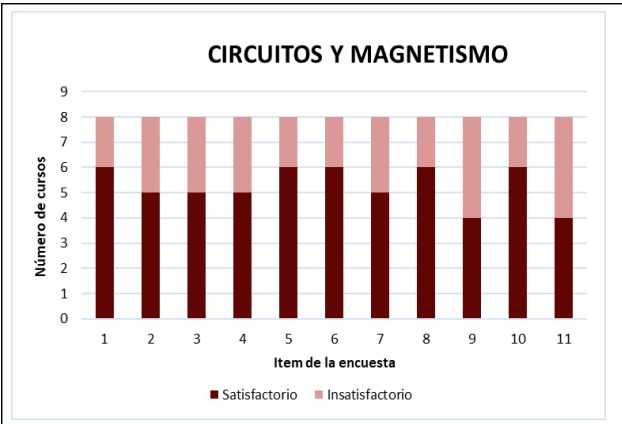
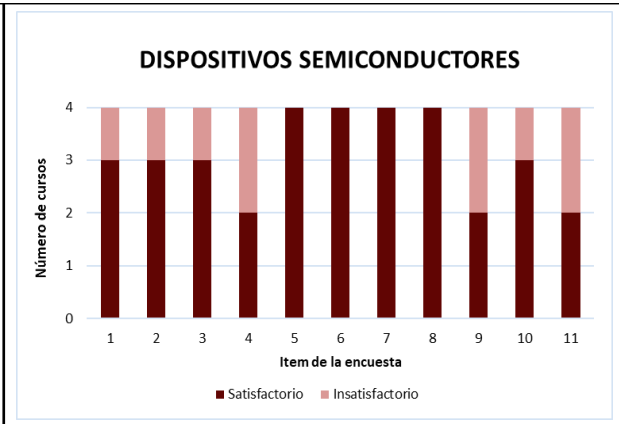
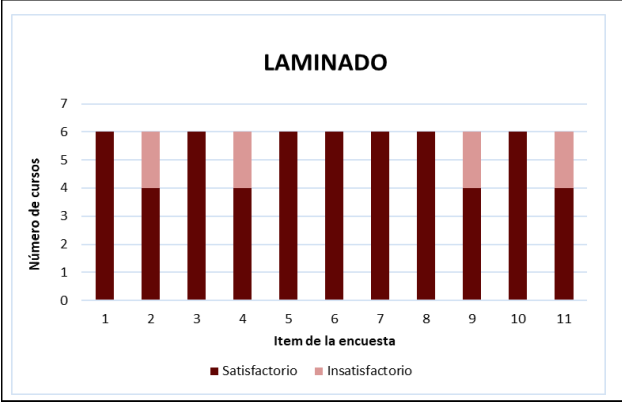
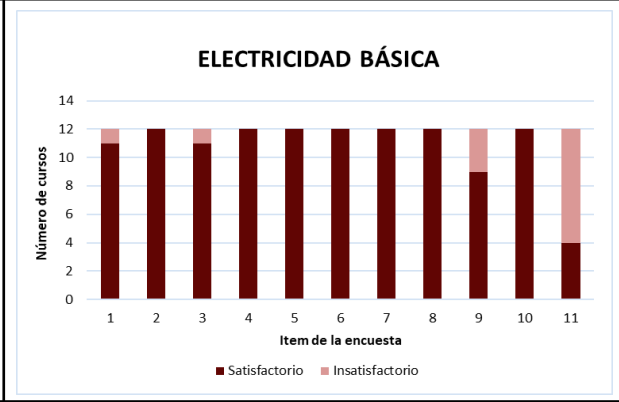
Las dimensiones que se muestran en el gráfico representan el 16 % de la variabilidad total, teniendo en cuenta que las dimensiones son 41, el porcentaje de variabilidad total es apropiado.

Tabla 5. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
CALC SOPLO	-0.26386310	0.007383775	-0.112692701	0.313395533	-0.388735300
CARA COQ	-0.32544605	-0.184727622	0.154232367	-0.381347845	-0.005950658
CARAC CYC	0.04887714	0.575803050	-0.599185111	-2.062.409.280	1.732.610.459
CIRC MAG	317.206.962	-1.178.650.573	-0.075671967	0.404752979	0.472094607
CLAS CHA	-0.23307163	0.103439474	-0.246155235	0.660767222	-0.580127621
⋮	⋮	⋮	⋮	⋮	⋮
TERM	-0.28439075	-0.056653357	-0.023717678	0.081814407	-0.261140419

Para identificar los ítems que causaron que en las formaciones identificadas mediante ACM con mayor asociación a la insatisfacción se procede a realizar el análisis individual de cada formación: Ver Anexo 5 5

Tabla 6. Análisis individual

 <p>CIRCUITOS Y MAGNETISMO</p> <table border="1"> <thead> <tr> <th>Item de la encuesta</th> <th>Satisfactorio</th> <th>Insatisfactorio</th> </tr> </thead> <tbody> <tr><td>1</td><td>6</td><td>2</td></tr> <tr><td>2</td><td>5</td><td>3</td></tr> <tr><td>3</td><td>5</td><td>3</td></tr> <tr><td>4</td><td>5</td><td>3</td></tr> <tr><td>5</td><td>6</td><td>2</td></tr> <tr><td>6</td><td>6</td><td>2</td></tr> <tr><td>7</td><td>5</td><td>3</td></tr> <tr><td>8</td><td>6</td><td>2</td></tr> <tr><td>9</td><td>4</td><td>4</td></tr> <tr><td>10</td><td>6</td><td>2</td></tr> <tr><td>11</td><td>4</td><td>4</td></tr> </tbody> </table>	Item de la encuesta	Satisfactorio	Insatisfactorio	1	6	2	2	5	3	3	5	3	4	5	3	5	6	2	6	6	2	7	5	3	8	6	2	9	4	4	10	6	2	11	4	4	 <p>DISPOSITIVOS SEMICONDUCTORES</p> <table border="1"> <thead> <tr> <th>Item de la encuesta</th> <th>Satisfactorio</th> <th>Insatisfactorio</th> </tr> </thead> <tbody> <tr><td>1</td><td>3</td><td>1</td></tr> <tr><td>2</td><td>3</td><td>1</td></tr> <tr><td>3</td><td>3</td><td>1</td></tr> <tr><td>4</td><td>2</td><td>2</td></tr> <tr><td>5</td><td>4</td><td>0</td></tr> <tr><td>6</td><td>4</td><td>0</td></tr> <tr><td>7</td><td>4</td><td>0</td></tr> <tr><td>8</td><td>4</td><td>0</td></tr> <tr><td>9</td><td>2</td><td>2</td></tr> <tr><td>10</td><td>3</td><td>1</td></tr> <tr><td>11</td><td>2</td><td>2</td></tr> </tbody> </table>	Item de la encuesta	Satisfactorio	Insatisfactorio	1	3	1	2	3	1	3	3	1	4	2	2	5	4	0	6	4	0	7	4	0	8	4	0	9	2	2	10	3	1	11	2	2
Item de la encuesta	Satisfactorio	Insatisfactorio																																																																							
1	6	2																																																																							
2	5	3																																																																							
3	5	3																																																																							
4	5	3																																																																							
5	6	2																																																																							
6	6	2																																																																							
7	5	3																																																																							
8	6	2																																																																							
9	4	4																																																																							
10	6	2																																																																							
11	4	4																																																																							
Item de la encuesta	Satisfactorio	Insatisfactorio																																																																							
1	3	1																																																																							
2	3	1																																																																							
3	3	1																																																																							
4	2	2																																																																							
5	4	0																																																																							
6	4	0																																																																							
7	4	0																																																																							
8	4	0																																																																							
9	2	2																																																																							
10	3	1																																																																							
11	2	2																																																																							
<p>Es el entrenamiento con índice de satisfacción general más bajo, inusualmente se encuentra insatisfacción con el formador (ítem 5 al 8), situación que no ocurre con los demás entrenamientos. A demás presenta baja satisfacción en el bloque de los ejes temáticos (ítem 1) al 4 y aún más en logística (ítem 9 al 11).</p>	<p>Los colaboradores se mostraron insatisfechos principalmente con el material entregado para el curso (ítem 4), la información previa a la formación (ítem 9) y la duración de la misma (ítem 11).</p>																																																																								
 <p>LAMINADO</p> <table border="1"> <thead> <tr> <th>Item de la encuesta</th> <th>Satisfactorio</th> <th>Insatisfactorio</th> </tr> </thead> <tbody> <tr><td>1</td><td>6</td><td>0</td></tr> <tr><td>2</td><td>4</td><td>2</td></tr> <tr><td>3</td><td>6</td><td>0</td></tr> <tr><td>4</td><td>4</td><td>2</td></tr> <tr><td>5</td><td>6</td><td>0</td></tr> <tr><td>6</td><td>6</td><td>0</td></tr> <tr><td>7</td><td>6</td><td>0</td></tr> <tr><td>8</td><td>6</td><td>0</td></tr> <tr><td>9</td><td>4</td><td>2</td></tr> <tr><td>10</td><td>6</td><td>0</td></tr> <tr><td>11</td><td>4</td><td>2</td></tr> </tbody> </table>	Item de la encuesta	Satisfactorio	Insatisfactorio	1	6	0	2	4	2	3	6	0	4	4	2	5	6	0	6	6	0	7	6	0	8	6	0	9	4	2	10	6	0	11	4	2	 <p>ELECTRICIDAD BÁSICA</p> <table border="1"> <thead> <tr> <th>Item de la encuesta</th> <th>Satisfactorio</th> <th>Insatisfactorio</th> </tr> </thead> <tbody> <tr><td>1</td><td>11</td><td>1</td></tr> <tr><td>2</td><td>12</td><td>0</td></tr> <tr><td>3</td><td>11</td><td>1</td></tr> <tr><td>4</td><td>12</td><td>0</td></tr> <tr><td>5</td><td>12</td><td>0</td></tr> <tr><td>6</td><td>12</td><td>0</td></tr> <tr><td>7</td><td>12</td><td>0</td></tr> <tr><td>8</td><td>12</td><td>0</td></tr> <tr><td>9</td><td>9</td><td>3</td></tr> <tr><td>10</td><td>12</td><td>0</td></tr> <tr><td>11</td><td>4</td><td>8</td></tr> </tbody> </table>	Item de la encuesta	Satisfactorio	Insatisfactorio	1	11	1	2	12	0	3	11	1	4	12	0	5	12	0	6	12	0	7	12	0	8	12	0	9	9	3	10	12	0	11	4	8
Item de la encuesta	Satisfactorio	Insatisfactorio																																																																							
1	6	0																																																																							
2	4	2																																																																							
3	6	0																																																																							
4	4	2																																																																							
5	6	0																																																																							
6	6	0																																																																							
7	6	0																																																																							
8	6	0																																																																							
9	4	2																																																																							
10	6	0																																																																							
11	4	2																																																																							
Item de la encuesta	Satisfactorio	Insatisfactorio																																																																							
1	11	1																																																																							
2	12	0																																																																							
3	11	1																																																																							
4	12	0																																																																							
5	12	0																																																																							
6	12	0																																																																							
7	12	0																																																																							
8	12	0																																																																							
9	9	3																																																																							
10	12	0																																																																							
11	4	8																																																																							
<p>Se vuelve a presentar insatisfacción con los ítems 4, 9 y 11. Al parecer son los puntos que más afectan la satisfacción para los entrenamientos específicos que presentan menor índice.</p>	<p>EL principal ítem que afectó el entrenamiento fue la duración de la formación (ítem 11). Se debe considerar incluir en la evaluación de satisfacción una aclaración sobre si este tiempo debió ser más largo o más corto.</p>																																																																								

4.1.2. ACM Satisfacción Vs Mes

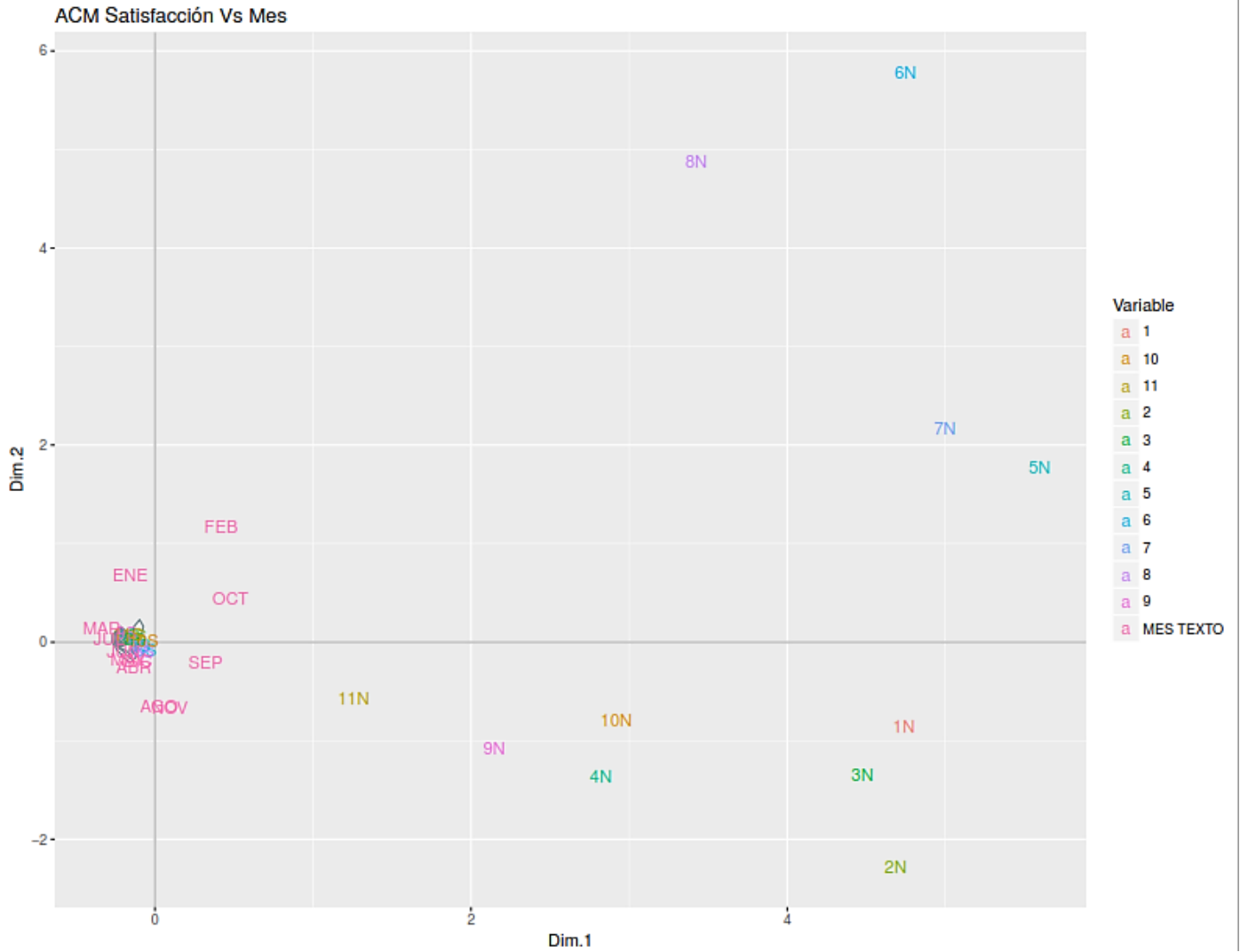


FIGURA 4: Satisfacción Vs Mes. Fuente El autor 2018

Gráficamente se observa que, en términos generales, la satisfacción del personal es alta pues se evidencia la concentración de los meses en torno a los ítems que resultaron satisfechos, sin embargo, aunque los meses no tienen asociación fuerte con la insatisfacción, los que tienen mayor asociación son febrero, octubre, septiembre, noviembre y agosto.

Tabla 7. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
3N	1,2332E+10	2,9004E+09	3,9420E+09	5,7096E+02	5,7745E-01
1N	1,1852E+10	1,0070E+09	5,8115E+09	3,8995E+05	6,8179E+09
2N	1,1591E+10	7,2152E+09	8,6692E-01	1,3227E+06	8,9225E-01
5N	1,1015E+10	2,8970E+09	1,2602E+10	5,8793E+05	1,5641E-01
7N	1,0988E+10	5,3926E+09	6,4731E+09	1,2398E+06	6,9892E-02
⋮	⋮	⋮	⋮	⋮	⋮
AGO	1,2843E-03	2,0438E+09	2,2459E+09	8,4131E+06	3,7112E+09

Como se observa en la tabla, las categorías que mayor contribución tienen a la varianza son los ítems que resultaron no satisfechos, entre los meses se destaca octubre que es el mes con mayor asociación a la insatisfacción, también febrero, marzo y julio.

Tabla 8. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.386530220	21,0834666	21,08347
dim 2	0.147607935	8,0513419	29,13481
dim 3	0.107539367	5,8657837	35,00059
dim 4	0.098693149	5,3832627	40,38385
⋮	⋮	⋮	⋮
dim 22	⋮	⋮	⋮

Las dimensiones que se muestran en el gráfico representan el 29 % de la variabilidad total, teniendo en cuenta que las dimensiones son 22, el porcentaje de variabilidad total es apropiado.

Tabla 9. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
6N	474.679.051	578.402.395	3.025.875.118	0.734962542	0.559463364
6S	-0.05884451	-0.07170278	-0.037510849	-0.009111106	-0.006935496
7N	499.724.600	216.343.305	-2.023.155.960	-0.848225376	-0.197988199
7S	-0.10410929	-0.04507152	0.042149083	0.017671362	0.004124754
8N	342.382.573	487.484.240	1.764.677.600	0.858096048	-0.435710372
⋮	⋮	⋮	⋮	⋮	⋮
SEP	0.32041546	-0.20795984	-1.335.903.623	-0.135607780	0.152820490

Para identificar los ítems que causaron que en las formaciones identificadas mediante ACM con mayor asociación a la insatisfacción se procede a realizar el análisis individual de cada mes: *Ver Anexo 6 5*

Tabla 10. Análisis individual

<p>OCTUBRE</p>	<p>FEBRERO</p>
<p>La mayor frecuencia está en que los colaboradores no estuvieron de acuerdo con el tiempo destinado a la capacitación (ítem 11). De igual manera se observa que los colaboradores consideran que la información entregada de manera previa a la formación no fue oportuna y/o apropiada (ítem 9) y que los materiales y recursos que acompañan la formación no cumplió con sus expectativas (ítem 4)</p>	<p>Los colaboradores se mostraron insatisfechos durante este mes con el bloque del instructor (ítem 5, 6, 7 y 8), la información previa a la formación (ítem 9) y la duración de la misma (ítem 11).</p>
<p>SEPTIEMBRE</p>	<p>NOVIEMBRE</p>
<p>En septiembre los bloques temática y logística son los que presentan mayor insatisfacción, en el bloque del instructor el ítem más afectado es la habilidad del formador para ayudar en la comprensión de los temas con ejemplos claros, remarcar lo importante y realizar conclusiones (ítem 7)</p>	<p>Se vuelve a presentar insatisfacción con los ítems 4, 9 y 11. Al parecer son los puntos que más afectan la satisfacción tanto mes a mes como específicamente en las formaciones identificadas con mayor índice de insatisfacción.</p>

Simultáneamente se observa que a lo largo del año los ítems 11, 9 y 4 mencionados anteriormente fueron los que obtuvieron mayor frecuencia de insatisfacción lo que indica que estos tres ítems son puntos débiles en el servicio prestado. Aunque el bloque del formador no tiene frecuencias altas en insatisfacción es importante resaltar que para este bloque los ítems más afectados son: la habilidad del formador para ayudar en la comprensión de los temas con ejemplos claros, remarcar lo importante y realizar conclusiones (ítem 7) y el uso de un lenguaje, ritmo de exposición y tono de voz apropiado (ítem 8).

4.2. APROBACIÓN

4.2.1. ACM Aprobación Vs Datos personales

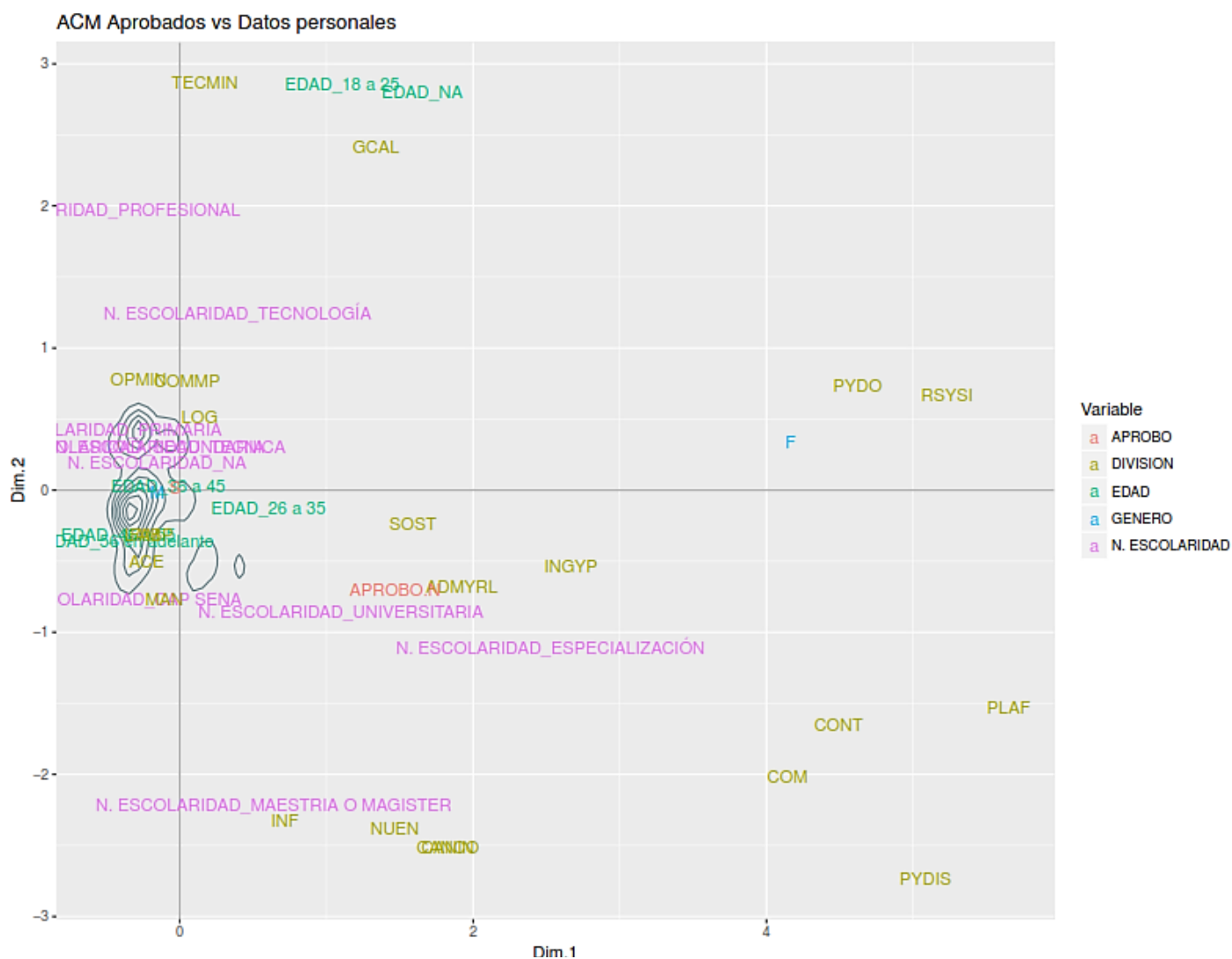


FIGURA 5: Aprobación Vs Datos personales. Fuente El autor 2018

Gráficamente se observa que la mayor concentración de categorías se presenta en torno a los colaboradores que aprobaron, sin embargo, la no aprobación se ve claramente asociada al área administrativa y de relaciones laborales, también a sostenibilidad y a ingeniería y proyectos.

En cuanto al nivel de escolaridad presenta mayor asociación con la no aprobación quienes tienen título profesional o especialización. Al parecer la edad no tiene asociación con la no aprobación, sin embargo, es prudente evaluar el rango de edad de 26 a 35 años que tiene asociación débil. El género tampoco revela asociación a la no aprobación.

Tabla 11. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
F	3,0902E+07	2,9131E-01	1,5522E+03	1,2583E-01	1,3301E+06
PYDO	1,23E+07	4,44E-01	5,21E+02	1,50E-01	2,08E+06
UNIVERSITARIA	8,0748E+06	7,0939E+08	3,3510E+06	5,7609E+09	5,1684E+04
ESPECIALIZACIÓN	7,5744E+06	2,1066E+08	1,1316E+07	2,1582E+10	2,7913E+05
18 a 25	2,7057E+06	2,6033E+09	1,9116E+06	8,4383E-01	8,7562E-02
APROBO.N	2,19E+06	7,28E-01	2,42E+06	3,80E+09	6,91E+05
⋮	⋮	⋮	⋮	⋮	⋮
COMMP	3,4184E+02	1,1787E-01	1,5693E+05	9,6130E-02	2,9398E+06

Categorías como el nivel académico universitario, especialización, la edad de 26 a 35 años, el área de Sostenibilidad identificadas en la gráfica como asociadas a no aprobación están entre las variables que más contribuyen a la varianza.

Tabla 12. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.40639777	5,4918617	5,491862
dim 2	0.28223289	3,813958	9,30582
dim 3	0.26669482	3,6039841	12,909804
dim 4	0.25236233	3,4103018	16,320106
⋮	⋮	⋮	⋮
dim 37	⋮	⋮	⋮

Las dimensiones que se muestran en el gráfico representan el 9% de la variabilidad total, aunque hay 37 dimensiones la variabilidad representada en el gráfico puede carecer concordancia con los datos reales.

Tabla 13. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
F	416.249.352	0.33679303	0.0238980458	-0.209305993	0.6731010570
M	-0.15652700	-0.01266481	-0.0008986655	0.007870773	-0.0253113885
18 a 25	110.629.743	285.972.629	0.7532935090	0.486851063	0.0001551194
26 a 35	0.60336238	-0.12357968	0.3595257469	-0.619349394	-0.4394856401
46 a 55	-0.42071824	-0.31906795	-0.0506559881	-0.104445220	0.5727722652
⋮	⋮	⋮	⋮	⋮	⋮
S	-0.03091199	0.01485644	0.0263151240	0.032085689	0.0135393902

Tabla 14. Análisis individual

<p>Aprobación Vs Edad</p>	<p>Aprobaciones Vs División</p>
<p>Aunque es evidente que la no aprobación es baja se presenta con mayor frecuencia entre los jóvenes de 26 a 35 años con 3,0% seguido de 36 a 45 años con 2,1%</p>	<p>Mediante el análisis por división se observa que las categorías que aparecen mediante ACM como asociadas a la reprobación realmente no tienen frecuencias altas, seguramente esto se debe a la poca varianza que explica el gráfico.</p>
<p>Aprobaciones Vs Genero</p>	<p>Aprobaciones Vs Nivel Académico</p>
<p>A pesar de que el análisis muestra que el género no está asociado a la no aprobación, se identifica que las mujeres tienen un índice de reprobación mayor al de los hombres con 7,5%.</p>	<p>El índice de no aprobación según el nivel académico de acuerdo al ACM indica que especialización con 4,27% y universitaria con 4,26% son los niveles con mayor índice de reprobación, aunque este es bajo respecto a la aprobación, seguido de tecnología con 4,03% y primaria con 3,22%. No se encuentran diferencias significativas entre estos 4 niveles.</p>

Tabla 15. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
2h	2,5214E+10	6,9036E+06	5,6073E+05	4,2067E+06	3,1340E+02
TAR	1,7772E+10	5,7796E+06	5,4322E+04	5,9810E+06	2,6908E+06
3h	7,37E+09	1,34E+07	5,13E+02	6,38E+06	1,06E+05
MYT	4,80E+09	3,73E+03	5,22E+04	1,18E+05	3,56E+04
APROBO.N	4,1901E+09	1,9753E+07	8,9158E+05	1,3842E+06	3,5457E+06
Lunes	4,0452E+09	6,8705E+05	6,8522E+05	2,9176E+06	2,9506E+06
⋮	⋮	⋮	⋮	⋮	⋮
JUN	2,32E-03	1,16E+06	4,48E+06	1,15E+07	9,91E+06

Categorías como la jornada de la tarde, tarde y noche, la intensidad horaria de 3 y 12 horas, el día lunes y el mes de abril, identificadas como asociadas a no aprobación están entre las variables que más contribuyen a la varianza.

Tabla 16. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.40933672	6,2020715	6,202071
dim 2	0.36712734	5,5625354	11,764607
dim 3	0.32517385	4,9268766	16,691483
dim 4	0.30387262	4,6041306	21,295614
⋮	⋮	⋮	⋮
dim 33	⋮	⋮	⋮

Las dimensiones que se muestran en el gráfico representan el 11,7% de la variabilidad total, aunque hay 33 dimensiones la variabilidad representada en el gráfico puede carecer de concordancia con los datos reales.

Tabla 17. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
APROBO . N	203.953.813	4.193.819.536	0.838527599	1.009.998.373	1.553.920.332
S	-0.04293289	-0.088281156	-0.017651257	-0.021260768	-0.032710488
12h	131.287.800	2.425.124.106	-0.862758028	1.213.362.037	-2.729.239.324
2h	265.634.525	-1.316.350.362	-0.353070488	-0.934852302	0.007756597
7h	129.260.567	5.999.518.527	3.668.234.638	-2.768.248.494	9.111.118.477
⋮	⋮	⋮	⋮	⋮	⋮
SEP	0.48419290	0.123786892	0.105623734	-1.546.910.340	0.596955683

Tabla 18. Análisis individual

<p>Aprobaciones Vs Intensidad horaria</p> <table border="1"> <thead> <tr> <th>Intensidad horaria</th> <th>No Aprobó (%)</th> <th>Aprobó (%)</th> </tr> </thead> <tbody> <tr><td>7</td><td>69,23</td><td>30,77</td></tr> <tr><td>3</td><td>52,17</td><td>47,83</td></tr> <tr><td>12</td><td>16,98</td><td>83,02</td></tr> <tr><td>6</td><td>0,00</td><td>100,00</td></tr> <tr><td>8</td><td>0,00</td><td>100,00</td></tr> <tr><td>16</td><td>0,00</td><td>100,00</td></tr> <tr><td>2</td><td>0,00</td><td>100,00</td></tr> <tr><td>1</td><td>0,00</td><td>100,00</td></tr> <tr><td>4</td><td>0,00</td><td>100,00</td></tr> <tr><td>9</td><td>0,00</td><td>100,00</td></tr> <tr><td>5</td><td>0,00</td><td>100,00</td></tr> <tr><td>1,5</td><td>0,00</td><td>100,00</td></tr> <tr><td>2,5</td><td>0,00</td><td>100,00</td></tr> </tbody> </table>	Intensidad horaria	No Aprobó (%)	Aprobó (%)	7	69,23	30,77	3	52,17	47,83	12	16,98	83,02	6	0,00	100,00	8	0,00	100,00	16	0,00	100,00	2	0,00	100,00	1	0,00	100,00	4	0,00	100,00	9	0,00	100,00	5	0,00	100,00	1,5	0,00	100,00	2,5	0,00	100,00	<p>Aprobaciones Vs Día semana</p> <table border="1"> <thead> <tr> <th>Día semana</th> <th>No aprobó (%)</th> <th>Aprobó (%)</th> </tr> </thead> <tbody> <tr><td>Lunes</td><td>4,92</td><td>95,08</td></tr> <tr><td>Miércoles</td><td>5,55</td><td>94,45</td></tr> <tr><td>Martes</td><td>0,00</td><td>100,00</td></tr> <tr><td>Jueves</td><td>0,00</td><td>100,00</td></tr> <tr><td>Viernes</td><td>0,00</td><td>100,00</td></tr> <tr><td>Sabado</td><td>0,00</td><td>100,00</td></tr> </tbody> </table>	Día semana	No aprobó (%)	Aprobó (%)	Lunes	4,92	95,08	Miércoles	5,55	94,45	Martes	0,00	100,00	Jueves	0,00	100,00	Viernes	0,00	100,00	Sabado	0,00	100,00
Intensidad horaria	No Aprobó (%)	Aprobó (%)																																																														
7	69,23	30,77																																																														
3	52,17	47,83																																																														
12	16,98	83,02																																																														
6	0,00	100,00																																																														
8	0,00	100,00																																																														
16	0,00	100,00																																																														
2	0,00	100,00																																																														
1	0,00	100,00																																																														
4	0,00	100,00																																																														
9	0,00	100,00																																																														
5	0,00	100,00																																																														
1,5	0,00	100,00																																																														
2,5	0,00	100,00																																																														
Día semana	No aprobó (%)	Aprobó (%)																																																														
Lunes	4,92	95,08																																																														
Miércoles	5,55	94,45																																																														
Martes	0,00	100,00																																																														
Jueves	0,00	100,00																																																														
Viernes	0,00	100,00																																																														
Sabado	0,00	100,00																																																														
<p>Tal como se identificó en ACM las intensidades horarias de 7, 3 y 12 horas en las formaciones son donde se presenta mayor índice de reprobación, el cual va de 16,98% a 69,23% que se considera representativo y el cual debe estar como prioridad en la generación de planes de acción.</p>	<p>Se presenta mayor índice de reprobación el día lunes con 4,92% lo cual concuerda con lo encontrado mediante ACM</p>																																																															
<p>Aprobaciones Vs Jornada</p> <table border="1"> <thead> <tr> <th>Jornada</th> <th>No aprobó (%)</th> <th>Aprobó (%)</th> </tr> </thead> <tbody> <tr><td>Tarde</td><td>14,84</td><td>85,16</td></tr> <tr><td>Tarde y Noche</td><td>12,05</td><td>87,95</td></tr> <tr><td>Mañana y Tarde</td><td>0,00</td><td>100,00</td></tr> <tr><td>Mañana</td><td>0,00</td><td>100,00</td></tr> <tr><td>Noche</td><td>0,00</td><td>100,00</td></tr> </tbody> </table>	Jornada	No aprobó (%)	Aprobó (%)	Tarde	14,84	85,16	Tarde y Noche	12,05	87,95	Mañana y Tarde	0,00	100,00	Mañana	0,00	100,00	Noche	0,00	100,00	<p>Aprobaciones Vs Mes</p> <table border="1"> <thead> <tr> <th>Mes</th> <th>No Aprobó (%)</th> <th>Aprobó (%)</th> </tr> </thead> <tbody> <tr><td>MARZO</td><td>7,28</td><td>92,72</td></tr> <tr><td>ENERO</td><td>4,55</td><td>95,45</td></tr> <tr><td>AGOSTO</td><td>0,00</td><td>100,00</td></tr> <tr><td>OCTUBRE</td><td>0,00</td><td>100,00</td></tr> <tr><td>FEBRERO</td><td>0,00</td><td>100,00</td></tr> <tr><td>NOVIEMBRE</td><td>0,00</td><td>100,00</td></tr> <tr><td>SEPTIEMBRE</td><td>0,00</td><td>100,00</td></tr> <tr><td>ABRIL</td><td>0,00</td><td>100,00</td></tr> <tr><td>JULIO</td><td>0,00</td><td>100,00</td></tr> <tr><td>JUNIO</td><td>0,00</td><td>100,00</td></tr> <tr><td>MAYO</td><td>0,00</td><td>100,00</td></tr> <tr><td>DICIEMBRE</td><td>0,00</td><td>100,00</td></tr> </tbody> </table>	Mes	No Aprobó (%)	Aprobó (%)	MARZO	7,28	92,72	ENERO	4,55	95,45	AGOSTO	0,00	100,00	OCTUBRE	0,00	100,00	FEBRERO	0,00	100,00	NOVIEMBRE	0,00	100,00	SEPTIEMBRE	0,00	100,00	ABRIL	0,00	100,00	JULIO	0,00	100,00	JUNIO	0,00	100,00	MAYO	0,00	100,00	DICIEMBRE	0,00	100,00						
Jornada	No aprobó (%)	Aprobó (%)																																																														
Tarde	14,84	85,16																																																														
Tarde y Noche	12,05	87,95																																																														
Mañana y Tarde	0,00	100,00																																																														
Mañana	0,00	100,00																																																														
Noche	0,00	100,00																																																														
Mes	No Aprobó (%)	Aprobó (%)																																																														
MARZO	7,28	92,72																																																														
ENERO	4,55	95,45																																																														
AGOSTO	0,00	100,00																																																														
OCTUBRE	0,00	100,00																																																														
FEBRERO	0,00	100,00																																																														
NOVIEMBRE	0,00	100,00																																																														
SEPTIEMBRE	0,00	100,00																																																														
ABRIL	0,00	100,00																																																														
JULIO	0,00	100,00																																																														
JUNIO	0,00	100,00																																																														
MAYO	0,00	100,00																																																														
DICIEMBRE	0,00	100,00																																																														
<p>Los entrenamientos donde se presenta mayor índice de reprobación suelen ofrecerse en las jornadas de la Tarde con 14,84% y Tarde-noche con 12,05%, datos que deben ser objeto de estudio para proponer planes de acción.</p>	<p>El mes donde se presenta mayor reprobación es en marzo con un índice de 7,28%, seguido de enero con 4,55%..</p>																																																															

4.3. ASISTENCIA

4.3.1. ACM Asistencia Vs Datos personales



FIGURA 7: Asistencia Vs Datos personales. Fuente El autor 2018

Gráficamente se observa que asociación fuerte en torno a la asistencia y la no asistencia, aparentemente se presentan altos índices de inasistencia en las edades de 26 años en adelante, las áreas de Logística, Compra De Materia Prima, Operaciones Mineras, Laminación, Acería y Fab. Primaria, con niveles académicos Técnica, SENA y universitarios y principalmente el género masculino.

Tabla 19. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
F	3,0937E+10	1,9570E-01	1,2117E+04	2,6741E-01	6,8395E+05
PYDO	1,3008E+10	2,8563E-01	2,3108E+04	1,5660E-02	7,9845E+05
UNIVERSITARIA	8,08E+09	7,74E+09	1,98E+06	2,51E+08	4,14E+06
ESPECIALIZACIÓN	7,8629E+09	1,3878E+09	1,3236E+07	1,2044E+09	8,9091E+06
CONT	7,7574E+09	2,1443E+09	1,3016E+05	4,7661E-02	2,4560E+06
18 a 25	3,02E+09	2,54E+10	3,26E+06	1,56E-02	1,78E+06
⋮	⋮	⋮	⋮	⋮	⋮
COMMP	5,22E-03	1,14E-01	5,63E+04	1,14E-01	1,96E+06

Categorías como el nivel académico universitario y SENA, las edades de 26 a 55 años, el género masculino, identificadas como asociadas a no aprobación están entre las variables que más contribuyen a la varianza en el factor 1, de igual manera con Laminación, Acería, Fab. Primaria en el factor 2.

Tabla 20. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.39445485	5,330471	5,330471
dim 2	0.27856088	3,7643363	9,094807
dim 3	0.26200399	3,5405945	12,635402
dim 4	0.25056083	3,3859571	16,021359
⋮	⋮	⋮	⋮
dim 37	⋮	⋮	⋮

Las dimensiones que se muestran en el gráfico representan el 9,09 % de la variabilidad total, aunque hay 37 dimensiones la variabilidad representada en el gráfico puede carecer de concordancia con los datos reales.

Tabla 21. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
ACE	-0.18058310	-0.418923114	0.104893143	0.20917359	-0.34448814
ADMYRL	216.715.064	0.188776805	-1.742.151.230	422.431.557	198.971.077
FABP	-0.16037570	-0.385656980	0.468654552	0.08377072	0.05910837
GCAL	135.860.411	2.318.301.943	1.720.725.353	0.14805915	0.29530418
INGYP	303.424.100	-0.689634730	-3.883.182.521	293.650.801	406.697.003
⋮	⋮	⋮	⋮	⋮	⋮
S	0.07440054	0.053565801	-0.094210071	-0.15244158	-0.06492362

Tabla 22. Análisis individual

<p>Asistencia Vs Edad</p>	<p>Asistencia Vs División</p>
<p>Se observa que a medida que aumenta la edad se presenta mayor inasistencia lo que concuerda con lo encontrado en ACM donde se presenta asociación en la edad de 26 años en adelante. Se debe considerar estudiar las causas que generan que se pierda el interés en la asistencia a medida que aumenta la edad de los colaboradores.</p>	<p>Se observa que las áreas identificadas mediante ACM como asociadas a la no asistencia efectivamente tienen frecuencias altas de inasistencia, sin embargo, hay otras áreas con frecuencias altas también que no se habían identificado y también deben ser foco de atención como: Sostenibilidad, Mantenimiento, Planeación distribución, Gestión Calidad, entre otras.</p>
<p>Asistencia Vs Genero</p>	<p>Asistencia Vs Nivel Académico</p>
<p>De acuerdo a ACM los hombres son quienes presentan inasistencia con mayor frecuencia</p>	<p>Respecto al nivel académico no se encuentran diferencias significativas entre los niveles que van de Técnica a Universitaria, todos presentan niveles altos de inasistencia (por encima del 20%)</p>

4.3.2. ACM Asistencia Vs Datos temporales



FIGURA 8: Asistencia Vs Datos temporales. Fuente El autor 2018

Gráficamente se observa que hay asociación fuerte en torno a la asistencia y la no asistencia, aparentemente se presentan altos índices de inasistencia en los meses de enero, febrero, marzo, octubre y agosto, la jornada mañana-tarde y tarde-noche, las formaciones de 8 y 16 joras de intensidad, en los días martes, miércoles, jueves, viernes y sábados

Tabla 23. Categorías con mayor contribución

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
2h	2,7896E+10	2,6300E-01	1,7211E-01	8,1567E+06	1,5127E+08
TAR	1,3151E+10	3,0433E+09	1,2995E+09	1,9272E+07	2,3219E-02
Lunes	5,5166E+09	2,6861E+09	3,1637E+09	2,6002E+02	9,0040E-01
8h	4,4815E+09	5,2882E-01	3,3796E+09	7,0434E+05	4,8743E-01
ASISTIO.N	3,9306E+09	3,6306E-01	1,9841E-01	6,6714E+05	3,8020E+08
MYT	3,6768E+09	5,5116E-02	2,6340E-01	8,8803E+02	3,9662E-02
⋮	⋮	⋮	⋮	⋮	⋮
Sábado	2,2444E-03	8,9667E-02	2,0104E-01	1,5198E+06	1,9113E+08

Categorías como el mes de febrero, la jornada mañana-tarde, las formaciones de 8 horas, identificadas como asociadas a no aprobación están entre las variables que más contribuyen a la varianza en el factor 1, de igual manera con los días martes y miércoles en el factor 2.

Tabla 24. Valores propios y porcentajes

	Eigen value	Percentage of variance	Cumulative percentage of variance
dim 1	0.41101648	6,2275224	6,227522
dim 2	0.32987187	4,9980587	11,225581
dim 3	0.32149855	4,8711902	16,096771
dim 4	0.29638433	4,4906717	20,587443
⋮	⋮	⋮	⋮
dim 33	⋮	⋮	⋮

Las dimensiones que se muestran en el gráfico representan el 11,22 % de la variabilidad total, aunque hay 33 dimensiones la variabilidad representada en el gráfico puede carecer de concordancia con los datos reales.

Tabla 25. Coordenadas

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
MAÑ	293.448.627	-0.10325092	-0.28673754	-1.429.065.484	-0.74312110
MYT	-0.29119592	-0.03194006	-0.06893167	0.003842947	-0.02499543
Jueves	-0.13789750	-0.66643961	0.19274620	0.061977529	0.03638753
Martes	-0.14317672	0.37210660	-0.07478771	0.114475519	-0.28274709
Sábado	0.05193993	0.29411306	-0.43476860	-1.147.733.677	-125.267.339
⋮	⋮	⋮	⋮	⋮	⋮
SEP	0.66415855	0.19544336	117.379.259	-0.999314743	182.044.849

Tabla 26. Análisis individual

<p>De acuerdo con ACM las intensidades horarias de 16 y 8 horas son las que mayor frecuencia de inasistencia presentan, acompañadas de 6, 1 y 3 horas.</p>	<p>Acorde a ACM se presenta mayor índice de inasistencia los días martes, miércoles, jueves, viernes y sábado. Se evidencia necesario generar planes de acción en temas de asistencia a formación.</p>
<p>Acorde a ACM la jornada mañana-tarde es donde mayor inasistencia se presenta, seguido de tarde-noche.</p>	<p>El mes donde se presenta mayor inasistencia es enero con un índice elevado de 39,23% seguido de abril con 34,15% y febrero con 31,63%. La inasistencia a lo largo del año siempre tuvo índices considerables, es necesario generar planes de acción para este ítem.</p>

5. Conclusiones

- Los colaboradores insatisfechos generalmente han asistido a formaciones en los meses de octubre, septiembre, noviembre y febrero principalmente a formaciones de Circuitos y magnetismo (durante octubre y noviembre), Dispositivos semiconductores (durante noviembre), Laminado (durante septiembre) y electricidad básica (de mayo a junio)
- En el bloque Temática los colaboradores consideran que el desarrollo de la formación promueve estrategias de innovación, práctica y aplicación del aprendizaje a la práctica (ítem 3), fue el ítem que presentó mayor índice de satisfacción. Mientras que los recursos que acompañan la formación no cumplieron con las expectativas de los colaboradores (ítem 4), es el que menor satisfacción presenta.

- En el bloque Instructor la motivación que el instructor ejerce hacia el grupo para que sea posible aclarar inquietudes e intercambiar experiencias e ideas en un ambiente agradable (ítem 6) fue el que presentó mayor índice de satisfacción. Mientras que la habilidad del formador para ayudar en la comprensión de los temas con ejemplos claros, remarcar lo importante y realizar conclusiones (ítem 7) y el uso de un lenguaje, ritmo de exposición y tono de voz apropiado por parte del instructor (ítem 8) presenta menor satisfacción.
- En el bloque logística fue donde se evidencio mayores índices de insatisfacción, el ítem 10 (Se respetaron y cumplieron las fechas, tiempos y horarios establecidos) presentó menor insatisfacción respecto a los demás ítems del bloque. Mientras el ítem 11 (los colaboradores no estuvieron de acuerdo con el tiempo destinado a la capacitación) fue el que mayor índice de insatisfacción tuvo para este bloque y a nivel global en la encuesta.
- Los colaboradores que mayor índice de reprobación tienen generalmente tienen edades entre 26 y 45 años, principalmente de las áreas de Planeación y Distribución, Responsabilidad Social y Relación Institucional, Compras, Supply Chain, con títulos académicos de especialización, universitarios y tecnólogos, en formaciones con intensidades horarias de 7, 3 y 12 horas, los días lunes y miércoles, en las jornadas de la tarde o tarde-noche.
- Los colaboradores que mayor inasistencia tuvieron generalmente son de las áreas de Logística, Compra de Materia Prima, Sostenibilidad, Fabricación Primaria, Mantenimiento, Planeación y Distribución, Laminación, Gestión de Calidad, Acería, Operaciones Mineras. La inasistencia aumenta a medida que aumenta la edad, usualmente son hombres, no hay distinción significativa entre el nivel académico. Suele haber mayor inasistencia a formaciones con intensidades de 16, 8, 6 y 1 hora, en los días de martes a sábado, durante las jornadas mañana-tarde y tarde-noche.

Referencias Bibliográficas

- Baquero Corrales, A. (2005), *Administración de recursos humanos (II parte)*, Editorial Universidad Estatal a Distancia EUED.
- Billorou, N., Pacheco, M. y Vargas, F. (2011), 'Guía para la evaluación de impacto de la formación', *Organización Internacional del Trabajo OIT/CINTERFOR*.
- De la Fuente García, D., Fernández, I., Fernández, N. G. y García, N. (2006), *Administración de empresas en ingeniería*, Universidad de Oviedo.
- de la Rosa de Sáa, S. (2012), Análisis estadístico comparativo de tres escalas de valoración: Likert, fuzzy-likert y fuzzy de respuesta libre, Master's thesis, Universidad de Oviedo.
- Díaz Monroy, L. G. y Morales Rivera, M. A. (2009), *Análisis de datos categóricos*, Editorial Universidad Nacional de Colombia.
- Flores Villalpando, R. et al. (2016), *Administración de Recursos Humanos*, Editorial Digital UNID.
- Franceschi Ramírez, C. y Hernández Pulido, M. (2007), 'Evaluación de modalidades de adiestramiento bajo el enfoque de Phillips (1996): Caso XEROX de Venezuela C.A.', *Univeridad Católica Andrés Bello*.
- García Parejo, S. y Sánchez, E. G. (2007), *Los recursos humanos aplicados a la gestión deportiva*, Vol. 658, INDE publicacones.
- López Dukmak, F. (2004), 'Visión internacional de la evaluación del impacto y retorno sobre la inversión en programas de capacitación en Chile, Argentina y Colombia', *Revista Facultad de Ciencias Económicas: Investigación y reflexión* **12**(1), 40-48.
- López Roldán, P. y Fachelli, S. (2015), 'Análisis de correspondencias - seminario de doctorado', *Diapositivas. Bellaterra: Universidad Autónoma de Barcelona*.

- Pardo Merino, A. y Ruiz Díaz, M. A. (2005), *OLC - Análisis de datos con SPSS 13 Base*, McGraw-Hill.
- Qian, X., Li, Z. y Zhou, Y. (2008), 'Modelo de formación y desarrollo a implementar en las universidades', *Biblioteca Virtual de derecho, economía y ciencias sociales* .
- Rutty, M. G. (2007), Evaluación de impacto en la capacitación de recursos humanos, PhD thesis, Facultad de Ciencias Económicas, Universidad de Buenos Aires.
- Supelano, J. M. (2014), 'Uso de la metodología roi para la medición de impacto y retorno sobre la inversión en programas de capacitación en la empresa tgt gamas s.a.s.', *Universidad Militar Nueva Granada* .
- Torres Ordóñez, J. L. (2005), 'Enfoques para la medición del impacto de la Gestión del Capital Humano en los resultados de negocio', *Pensamiento & Gestión* (No. 18), 151-176.

Anexo 1

MODELO DE KIRKPATRICK

Para Kirkpatrick la razón central para evaluar un programa es comprobar su efectividad. Expresa que existen tres razones para evaluar: a.- Justificar la existencia del departamento de formación mostrando como contribuye a los objetivos de la organización. b.- Para continuar o discontinuar un programa y c.- Para obtener información sobre cómo mejorar futuros programas.

Presenta un modelo de evaluación de la formación (entrenamiento) en cuatro niveles. Los cuatro niveles representan una secuencia de formas de evaluar programas. Cada nivel es importante y tiene impacto en el siguiente nivel. Advierte que ninguno de los niveles debería ser omitido para ir al nivel que el entrenador considera más importante. (Rutty 2007)

• Nivel 1: Reacción:

La evaluación en este nivel mide como los participantes reaccionan frente al programa (satisfacción). Si los participantes no reaccionan satisfactoriamente al programa, es probable que no estén motivados para aprender. Una reacción positiva no asegura aprendizaje, pero una reacción negativa reduce que ocurra el aprendizaje. Por eso es importante saber cuál es la satisfacción del asistente a los programas de formación. Las razones para medir la reacción son que:

- ◊ Genera un retorno que permite evaluar el programa y mejorarlo,
- ◊ Le dice a los entrenados que los capacitadores están para ayudarlos y que esta información les permite hacer su trabajo mejor,
- ◊ Genera información cuantitativa para ofrecerla a los jefes u otros entrenadores, y
- ◊ Provee información cuantitativa que puede ser usada para fijar estándares para evaluar el desempeño de futuros programas.

• Nivel 2: Aprendizaje

El aprendizaje es entendido como la medida en la cual el participante cambia actitudes, mejora conocimientos e incrementa habilidades como resultado de asistir al programa de formación. Evaluar el aprendizaje es importante, porque sin aprendizaje no ocurren cambios en la conducta. Las razones de falta de cambios en la conducta pueden deberse a un clima desfavorable de trabajo.

El incremento de conocimiento es fácil de medir, para medir actitudes se puede utilizar una prueba escrita, para evaluar habilidades se necesita una prueba de desempeño. Se recomienda la comparación antes y después y se sugiere la utilización de grupo control si es practicable.

• Nivel 3: Comportamiento

El comportamiento puede ser definido como la medida en la cual se produce un cambio en el comportamiento a causa de que los participantes asisten al programa de entrenamiento. Este nivel mide la transferencia del aprendizaje. Los cambios en el comportamiento en el trabajo ocurridos porque los capacitados asistieron al programa de formación. Es imposible predecir cuándo se producirá el cambio esperado, es importante procurar ayuda al capacitado para aplicar lo aprendido en el trabajo. Recomienda el uso grupo de control si es practicable, dejar pasar tiempo para que se produzca un cambio en la conducta (que varía según los cursos, pueden ser 6 meses), evaluar antes y después del entrenamiento, entrevistar a los entrenados, superiores inmediatos, sus subordinados y otros que a menudo observen su comportamiento, los aspectos sobre los que se indaga son:

- ◊ Si aplicó lo aprendido a su tarea
- ◊ Si el empleado mejoró su performance en la competencia involucrada
- ◊ Si lo aprendido y aplicado permaneció en el tiempo contribuyendo a los resultados del negocio

- **Nivel 4: Resultados**

Los resultados son los resultados finales que ocurren a causa de que los participantes asistieron al programa. Los resultados finales pueden incluir incremento en la producción, mejoramiento de la calidad, baja de costos, reducción de la frecuencia o severidad de accidentes, incremento de ventas, reducción de rotación y más altos beneficios. Los objetivos finales de la formación tienen necesariamente que ser redactados en esos mismos términos.

Desde esta perspectiva, es muy difícil, si no imposible, medir los resultados finales de programas como liderazgo, comunicación, motivación, management, empowerment, toma de decisiones o gerenciamiento del cambio. Se puede fijar y evaluar conductas deseadas, pero los resultados finales tienen que ser medidos en términos de mejoramiento moral u otros términos no financieros.

METODO ROI

- **Nivel 5: ROI**

Se mide el efecto de la formación en la productividad. Se busca medir en porcentaje y en moneda la incidencia de la formación en las ganancias del negocio (ROI). Si la formación se midiera sola se produciría un sobredimensionamiento de la influencia de la misma, por lo tanto, se analiza la contribución de la formación junto con los otros factores que determinan el negocio. Los criterios valorativos son los que definen los jefes subjetivamente a través de lo que pudiera concebirse como “juicio de expertos”. De este análisis se obtiene el retorno de la inversión (ROI) que es una medida monetaria ponderada sobre un juicio de valor sui géneris realizada por un juez (el jefe) en este caso concebido como “experto”.

Anexo 2

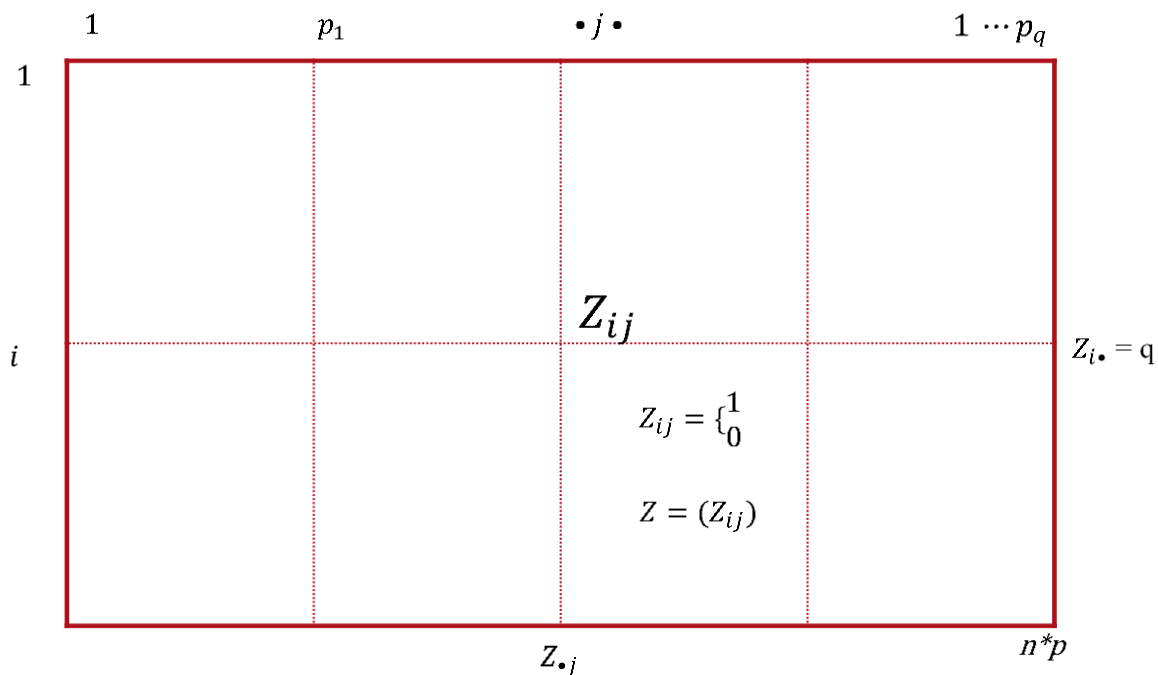


FIGURA 9: Tabla disyuntiva. Fuente Universidad de Costa Rica

Donde:

$n = \# \text{ individuos}$

$q = \# \text{ variables cualitativas}$

$p_k = \# \text{ categorías de la variable } k$

$$p = \sum_{k=1}^q p_k = \text{total de columnas}$$

Márgenes:

$Z_{i.} = q$ pues hay q 1 en la fila i

$Z_{.j} = \# \text{ individuos que tienen la categoría } j$

$$\sum_j \sum_i Z_{ij} = nq$$

(Universidad de Costa Rica)

Tabla 27. $Z =$ Matriz de disyuntiva o lógica binaria asociada a la matriz de datos

	M	F	Boy	Cun	Met	Qui	MB	B	M	A	MA
1	1	0	1	0	0	0	0	0	0	1	0
2	0	1	0	1	0	0	0	1	0	0	0
3	0	1	0	0	1	0	0	0	0	0	1
4	1	0	0	0	0	1	0	1	0	0	0
⋮ n	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Anexo 3

Tabla 28. $B=$ Matriz o Tabla de Contingencia de Burt, $B=Z^2Z$, es la que resulta de todas las posibles tablas de contingencia de las k variables.

	M	F	Boy	Cun	Met	Qui	MB	B	M	A	MA
M	140	0	40	35	42	23	15	29	46	37	13
F	0	123	38	28	37	20	8	32	51	18	14
Boy	40	38	78	0	0	0	6	18	29	16	9
Cun	35	28	0	63	0	0	4	12	24	15	8
Met	42	37	0	0	79	0	8	15	32	17	7
Qui	23	40	0	0	0	43	5	16	12	7	3
MB	15	8	6	4	8	5	23	0	0	0	0
B	29	32	18	12	15	16	0	61	0	0	0
M	46	51	29	24	32	12	0	0	97	0	0
A	37	18	16	15	17	7	0	0	0	55	0
MA	13	14	9	8	7	3	0	0	0	0	27

Anexo 4

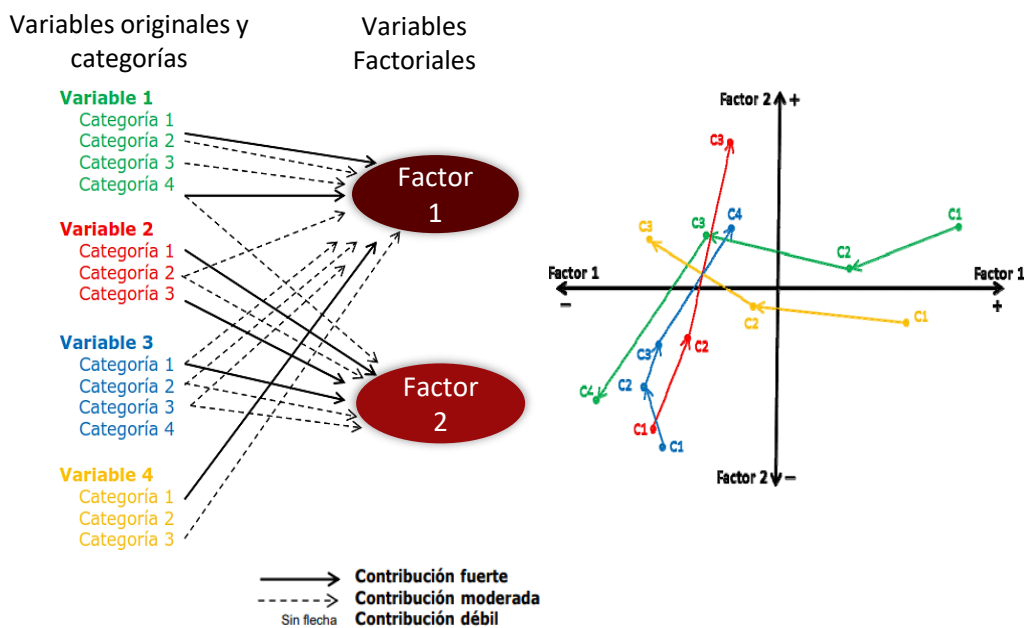


FIGURA 10: Representación gráfica ACM. Fuente López Roldan y Fachelli 2015

Anexo 5

Tabla 29. Satisfacción Vs Formación

	TEMÁTICAS											INSTRUCTOR						LOGÍSTICA						Total		Total	
	1		2		3		4		5		6		7		8		9		10		11		Sat	Insat	Sat	Insat	
	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	
CALCULO DE SOPLO BALANCE DE MASA	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	5	1	65	1	98%	2%	
CARACTERISTICAS DEL COQUE	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11	0	100%	0%	
CARACTERIZACION DEL CARBON Y CLASIFICACION	2	0	2	0	2	0	1	1	2	0	2	0	2	0	2	0	2	0	2	0	2	0	21	1	95%	5%	
CIRCUITOS Y MAGNETISMO	6	2	5	3	5	3	5	3	6	2	6	2	5	3	6	2	4	4	6	2	4	4	58	30	66%	34%	
CLASIFICACION DE CHATARRA	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	3	1	43	1	98%	2%	
CONVERTIDORES	8	0	8	0	8	0	7	1	8	0	8	0	8	0	8	0	7	1	8	0	8	0	86	2	98%	2%	
DESGASIFICACION EN HORNO CUCHARA	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	55	0	100%	0%	
DISPOSITIVOS SEMICONDUCTORES	3	1	3	1	3	1	2	2	4	0	4	0	4	0	4	0	2	2	3	1	2	2	34	10	77%	23%	
ELECTRICIDAD BASICA	11	1	12	0	11	1	12	0	12	0	12	0	12	0	12	0	9	3	12	0	4	8	119	13	90%	10%	
ESCORIAS EN AFINO SECUNDARIO	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	3	0	33	0	100%	0%	
FORMADOR DE FORMADORES	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11	0	100%	0%	
FUNDAMENTOS DE CALCINACION Y MATERIAS PRIMAS	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	22	0	100%	0%	
FUNDAMENTOS DE SIDERURGIA	48	1	49	0	47	2	45	4	47	2	48	1	47	2	46	3	47	2	49	0	44	5	517	22	96%	4%	
FUNDAMENTOS DEL ALTO HORNO	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	66	0	100%	0%	
GUIADO	3	1	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	3	1	4	0	42	2	95%	5%	
HORNOS VERTICALES	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	22	0	100%	0%	
LAMINADO	6	0	4	2	6	0	4	2	6	0	6	0	6	0	6	0	4	2	6	0	4	2	58	8	88%	12%	
MANEJO DE SUBPRODUCTOS Y GAS DE COQUE BRUTO	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11	0	100%	0%	
MATEMATICAS BASICAS	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	14	1	164	1	99%	1%	
MATERIAS PRIMAS	4	0	4	0	4	0	3	1	4	0	4	0	4	0	4	0	4	0	4	0	4	0	43	1	98%	2%	
METALURGIA SECUNDARIA	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	5	0	55	0	100%	0%	
METROLOGIA APLICADA	24	0	24	0	24	0	24	0	24	0	24	0	24	0	24	0	23	1	24	0	23	1	262	2	99%	1%	
MUESTREO DE CARBONES	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11	0	100%	0%	
MUESTREO DE MATERIAS PRIMAS Y CONSERVACION DE LA MUESTRA	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	1	0	10	1	91%	9%	
NIVELAR	11	0	11	0	11	0	11	0	11	0	11	0	11	0	11	0	9	2	11	0	11	0	119	2	98%	2%	
OPERACION ALTO HORNO	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	6	0	66	0	100%	0%	
PROCESO SINTER	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	11	0	100%	0%	
QUIMICA BASICA	8	0	8	0	8	0	8	0	8	0	8	0	8	0	8	0	6	2	7	1	6	2	83	5	94%	6%	
RECEPCION, ALMACENAMIENTO, CORTE, CARGUE Y CALENTAMIENTO DE PALANQUILLA	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	4	0	44	0	100%	0%	
SITUACIONES DE OPERACION ESPECIALES ALTO HORNO	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	2	0	22	0	100%	0%	
TERMODINAMICA	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	9	0	8	1	98	1	99%	1%	
TORNILLERIA	30	0	30	0	30	0	30	0	30	0	30	0	30	0	30	0	30	0	30	0	29	1	329	1	100%	0%	
TOTAL	239	6	239	6	238	7	231	14	241	4	242	3	240	5	240	5	226	19	239	6	216	29			2695		

Se puede observar que el porcentaje de satisfacción es alto en todos los entrenamientos, los entrenamientos con mayor porcentaje de insatisfacción son: Circuitos y magnetismo 34 %, dispositivos semiconductores 23 %, Laminado con 12 % y Electricidad básica 10%, comprobando los datos obtenidos por ACM. Se detectan falencias con la satisfacción de los entrenamientos dela ruta de mantenimiento eléctrico.

Anexo 6

Tabla 30. Satisfacción Vs Mes

	TEMÁTICAS						INSTRUCTOR						LOGÍSTICA						Total		Total					
	1		2		3		4		5		6		7		8		9		10		11		Total	Total		
	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat	Sat	Insat		
ENERO	9	0	9	0	9	0	8	1	9	0	9	0	9	0	8	1	9	0	9	0	9	0	97	2	98%	2%
FEBRERO	19	1	20	0	19	1	19	1	18	2	19	1	18	2	18	2	18	2	20	0	16	4	204	16	93%	7%
MARZO	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	22	0	21	1	241	1	100%	0%
ABRIL	16	0	16	0	15	1	14	2	16	0	16	0	16	0	16	0	15	1	16	0	16	0	172	4	98%	2%
MAYO	21	0	21	0	20	1	21	0	21	0	21	0	21	0	21	0	20	1	21	0	16	5	224	7	97%	3%
JUNIO	27	1	28	0	28	0	28	0	28	0	28	0	28	0	28	0	27	1	28	0	22	6	300	8	97%	3%
JULIO	27	0	27	0	27	0	27	0	27	0	27	0	27	0	27	0	25	2	27	0	26	1	294	3	99%	1%
AGOSTO	21	0	19	2	21	0	18	3	21	0	21	0	21	0	21	0	19	2	20	1	19	2	221	10	96%	4%
SEPTIEMBRE	18	1	18	1	18	1	17	2	18	1	19	0	17	2	19	0	17	2	18	1	17	2	196	13	94%	6%
OCTUBRE	25	1	24	2	24	2	23	3	25	1	24	2	25	1	24	2	23	3	24	2	22	4	263	23	92%	8%
NOVIEMBRE	21	1	21	1	21	1	20	2	22	0	22	0	22	0	22	0	19	3	21	1	18	4	229	13	95%	5%
DICIEMBRE	13	1	14	0	14	0	14	0	14	0	14	0	14	0	14	0	12	2	13	1	14	0	150	4	97%	3%
TOTAL	239	6	239	6	238	7	231	14	241	4	242	3	240	5	240	5	226	19	239	6	216	29	2695			

Los ítems con mayor grado de satisfacción en el año tienen que ver con el bloque dedicado al instructor donde los colaboradores resaltan su dominio del tema y manejo del grupo de trabajo (ítem 5) y la motivación que éste ejerce hacia el grupo para que sea posible aclarar inquietudes e intercambiar experiencias e ideas en un ambiente agradable (ítem 6). Se puede observar que el mes donde mayor insatisfacción se presentó fue en octubre seguido de febrero, septiembre y noviembre como se había identificado en ACM.



EFFECTO DE LA LISTA DE PALABRAS DE LA “SELF-COMPLEXITY TASK” EN LOS PUNTAJES DE “SELF-COMPLEXITY” EN ESTUDIANTES UNIVERSITARIOS

Especialización en Estadística

LORENA SANABRIA PÉREZ^{1,a}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

Este trabajo presenta la réplica de uno de los seis sub-estudios adelantados por Woolfolk et al (1995) para la evaluación de la validez de constructo de la “Self-Complexity Task”. El objetivo es determinar el efecto que tiene la lista de palabras que conforma la tarea en las puntuaciones finales que se obtienen al aplicarla. Para esto se diseñaron tres tareas, una con palabras predominantemente negativas, una predominantemente positiva y una neutra. También se analizó el efecto del sexo y los niveles de autoestima en las puntuaciones de “Self-Complexity”. Se realizó un análisis de varianza para grupos desbalanceados y se halló que ninguno de los tres factores tiene un efecto significativo sobre las puntuaciones a diferencia de lo encontrado por Woolfolk et al (1995). Los alcances de este hallazgo se revisan en la discusión.

Palabras clave: “Self-Complexity Task”, ANOVA, Sexo, Auto-estima, Valencia Afectiva.

Abstract

This work presents a replication of one Woolfolk’s et al (1995) six sub-studies advanced for the evaluation of the Self-Complexity Task’s construct validity. The objective is to determine the effect of the words list that make up the task in the final scores obtained when applying it. For this, three tasks were designed, one with predominantly negative words, one predominantly positive and one neutral. The effect of sex and self-esteem levels on “Self-Complexity” scores were also analyzed. An analysis of variance was performed for unbalanced groups and it was found that none of the three factors had a significant effect on the scores, unlike what was found by Woolfolk et al. The scope of this finding is reviewed in the conclusions.

Key words: “Self-Complexity Task”, ANOVA, Sexo, Auto-estima, Valencia Afectiva.

1. Introducción

“Self-complexity” (SC), el concepto creado por Linville (1985), implica que cada persona tiene una manera de estructurar la noción que tiene de sí mismo, y que esta estructura determina la susceptibilidad a verse afectado por los eventos estresantes del cotidiano. Para Linville (1985), una persona con baja complejidad, lo que significa que cuenta con menos roles significativos muy similares entre sí, tenderá a presentar mayores niveles de estrés y de enfermedades físicas derivadas del estrés, frente a las personas con alta complejidad, es decir, quienes tienen una descripción más diversificada de sí mismos. Para determinar el nivel de complejidad

^aEstudiante de Especialización en Estadística. E-mail: loresan288@hotmail.com

de cada individuo Linville creó la “Self-Complexity Task”.

Varios estudios siguieron la línea teórica y metodológica de Linville (1985, 1987), pero no todos encontraron los mismos resultados. Una explicación frecuente para estas diferencias se fundamentó en las condiciones del instrumento. Desde la estructura de la tarea hasta el sistema de cálculo han sido revisados y evaluados con el objetivo de encontrar la más precisa estrategia de medición, sin embargo, aún no se han unificado criterios.

Lo que se propone en este estudio es la reproducción de uno de los seis sub-estudios adelantados por Woolfolk et al (1995) para revisar las condiciones psicométricas de la “Self-Complexity Task” asumiendo esto como un primer paso que permitirá decidir a mediano plazo, la línea teórica a seguir para construir una versión en castellano de esta tarea y continuar con el estudio de este constructo.

2. Referente Conceptual

La “self-complexity task” fue el instrumento ideado por Linville (1987) para determinar el nivel de complejidad que del concepto de sí mismos tenían los individuos. Esta presentaba a los participantes 33 tarjetas cada una de las cuales contenía una característica o adjetivo. Lo que se les solicitaba a las personas es que pensarán en los roles significativos en su vida y los describieran usando las tarjetas, en la ejecución podían omitir el uso de tarjetas y repetir las, si lo consideraban conveniente, en los diferentes roles. Los datos eran procesados con el estadístico H , el cual computa las combinaciones de características presentes entre los roles. H se calcula con la siguiente fórmula:

$$H = \log_2 n - \frac{(\sum in_i \log_2 n_i)}{n} \quad (1)$$

En donde n es el número de características usada en la tarea y n_i es el número de elementos que cuenta con la misma combinación entre los roles definidos por la persona.

Linville (1987) obtuvo resultados que le permitieron confirmar su hipótesis y varios estudios posteriores le han respaldado (Niedenthal, Setterlund y Wherry 1992, Sakaki 2004, Heath 2011), sin embargo, ha habido otra porción de investigaciones que han mostrado datos inconsistentes y por tanto conclusiones diferentes acerca del rol de la alta complejidad como factor protector (Campbell, Chew y Scratchley 1991, Fankhauser 1991, McConnell, Strain, Brown y Rydell 2009).

Como consecuencia de estas diferencias se han desarrollado diferentes tipos de conceptualizaciones derivadas del postulado original de Linville (1987) las cuales abordan modificaciones en el esquema de cálculo o ajustes en la tarea, que conllevan a otras formas de observar el fenómeno.

Woolfolk, Novalany, Gara, Allen y Polino (1995) consideraron que uno de los problemas de la tarea propuesta por Linville (1987) fue el hecho de contar con más adjetivos positivos que negativos y de esta manera, las personas con depresión, tendencia al estrés u otros indicadores de malestar, no tenían las suficientes características para describirse, lo que los mostraría con bajos puntajes de complejidad.

Como resultado Woolfolk et al (1995) recomendaron para la “Self-Complexity Task” aumentar el número de características usadas a 44, equilibrando el número de acuerdo a su valencia y calcular separadamente la complejidad negativa y la complejidad positiva. Estos cambios no solo tienen efectos de forma, sino también en la manera de conceptualizar el fenómeno, por este motivo es de gran importancia, de cara a crear una versión en castellano de la “Self-Complexity Task” determinar la línea teórica que explica mejor el fenómeno.

Así las cosas, en este trabajo se seguirá el procedimiento planteado por Woolfolk, Novalany, Gara, Allen y Polino (1995) para determinar si la cantidad de palabras positivas o negativas en la lista de características usada en la tarea afecta los puntajes de “Self-Complexity”.

Análisis de varianza

Woolfolk et al (1995) usaron como técnica estadística el análisis de varianza (“Analysis of variance” - ANOVA) de tres vías. Esto significa que se observó el efecto de tres factores, el tipo de lista, el sexo y el nivel de auto-estima sobre la variable respuesta, que en este caso es el puntaje de SC y junto con esto, el efecto que puedan tener las interacciones entre factores, por lo que existen también fuentes de variación de segundo orden (interacción de dos factores) y de tercer orden (interacción entre los tres factores).

El modelo lineal que representa de manera general este tipo de interacciones es:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{ijk} \quad (2)$$

Sin embargo, se realiza una prueba de hipótesis para cada fuente de variación, en la que la hipótesis nula indica la igualdad en las medias y la hipótesis alterna que al menos dos medias en cada fuente de variación son diferentes (De la Garza García, Morales Serrano y González Cavazos 2013).

El estadístico de prueba en un análisis de varianza es la prueba F , y su valor se calcula con el cociente entre el cuadrado medio de cada fuente de variación y el cuadrado medio del residual. Los grados de libertad, en el caso de las fuentes de variación primaria, son el número de niveles que tiene el factor menos uno, para las fuentes secundarias y terciarias la multiplicación del número de niveles menos 1 de cada fuente de variación y para el residual la multiplicación de los niveles de cada factor y el número de sujetos que se encuentren en un nivel de cada factor menos uno (Aparicio, Martínez y Morales 2004, Valiente y Tejedor 2010).

La hipótesis nula se rechaza si el valor de F calculado es superior del F teórico y el p-valor es menor que el α propuesto (Valiente y Tejedor 2010).

La ejecución de un análisis de varianza requiere que se confirme que la distribución de la variable respuesta es normal y que existe homocedasticidad, sin embargo, es posible resolver el hecho de no contar con estas condiciones acudiendo, por ejemplo, a una transformación de la variable respuesta (Faraway 2014). Otros como Valiente & Tejedor (2010) afirman que el ANOVA presenta robustez frente al no cumplimiento de estos supuestos.

De otra parte, el que los grupos estén balanceados es otra condición importante en el análisis de varianza, pero existen procedimientos que permiten calcular los medios cuadrados de manera alternativa, sin que se cuente con el mismo número de elementos para cada factor de variación, estos son tipo I, llamado secuencial; tipo II, parcialmente secuencial y tipo III, marginal u ortogonal. El primero es indicado para casos desbalanceados en donde el orden de los factores es relevante, el segundo cuando no se requiere observar la interacción entre varios factores y el tercero cuando requiere observarse los efectos principales en presencia de las interacciones (Monleon-Getino 2017).

3. Diseño Metodológico

Población y Muestra

Participaron en el estudio 139 estudiantes (hombres y mujeres) de los 196 matriculados en la Licenciatura en Matemáticas y Estadística de la Universidad Pedagógica y Tecnológica de Colombia de la Seccional Duitama.

De estos 139 estudiantes, 47 elegidos de acuerdo al puntaje obtenido en la escala RSE formaron parte de la sub-muestra que resolvió la “Self-Complexity Task” en la segunda etapa de recolección

Instrumentos

Likert Frecuencia y Valencia: Es un instrumento creado para este estudio siguiendo las indicaciones de Woolfolk et al (1995), en el que se solicitó a las personas que calificaran 60 características en dos escalas tipo Likert, la primera de ellas con cinco opciones de respuesta en un diferencial semántico (siempre - nunca) para evaluar la frecuencia con que usarían cada palabra para describirse a sí mismos; y la segunda con siete opciones de respuesta, que van desde “Muy negativa” hasta “Muy positiva”, con la que se debía valorar la valencia del descriptor.

RSE – Rosenberg Self-Esteem (Rosenberg, 1965; adaptación de Atienza, Balaguer, & Moreno, 2000): Es un cuestionario de diez ítems (cinco directos y cinco inversos) diseñado para evaluar los sentimientos actuales de las personas relacionados consigo mismo. Para cada una de las frases que presenta la persona debe marcar una opción de las 4 posibles en escala Likert que va desde “Muy de acuerdo” hasta “Muy en desacuerdo”. Se obtienen puntajes brutos que oscilan entre 10 y 40 para cada persona. La versión en castellano cuenta con una fiabilidad de 0.80 y una consistencia interna de 0.76.

Self-Complexity Task (Linville 1987, Woolfolk, Novalany, Gara, Allen y Polino 1995): Tres adaptaciones de la “Self-Complexity Task” configuradas para este estudio fueron usadas. Las tres contaron con 33 palabras, como la versión original de (Linville 1987), pero se tuvo en cuenta para su construcción la valencia afectiva de las palabras, por lo que la primera lista estuvo conformada por 22 palabras negativas y 11 positivas, la segunda lista por 22 palabras positivas y 11 negativas y la tercera por 17 palabras positivas y 16 palabras negativas. Las instrucciones entregadas fueron las mismas que las de la tarea original de (Linville 1987). Los datos obtenidos fueron procesados con el estadístico H (ver ecuación 1). Los puntajes posibles se encuentran en el intervalo comprendido entre 0 y 5.04 que es el valor correspondiente al $\log_2 33$. Detalles acerca de la construcción de las listas se encuentran en el procedimiento.

Procedimiento

Pre-Test

Siguiendo lo propuesto por Woolfolk et Al (1995) se realizó una primera medición con el RSE para determinar las personas más aptas para participar en la segunda etapa quienes debían provenir del cuartil más bajo y más alto de acuerdo a las puntuaciones obtenidas en esta escala. La recolección de estos datos se realizó en línea a través de un enlace enviado a los correos de los estudiantes y otra se recogió en papel. Para todos los casos se solicitó a los participantes que firmaran un consentimiento informado para el tratamiento de sus datos.

En cuanto a las puntuaciones, para las mujeres el primer cuartil osciló entre los 21 y los 28 puntos y el cuarto cuartil entre los 35 y los 40 puntos. Para los hombres el primer cuartil empezó en los 22 puntos y terminó en los 29, el cuarto cuartil tuvo puntajes iguales al de las mujeres. La distribución de puntajes se observa en la gráfica 1.

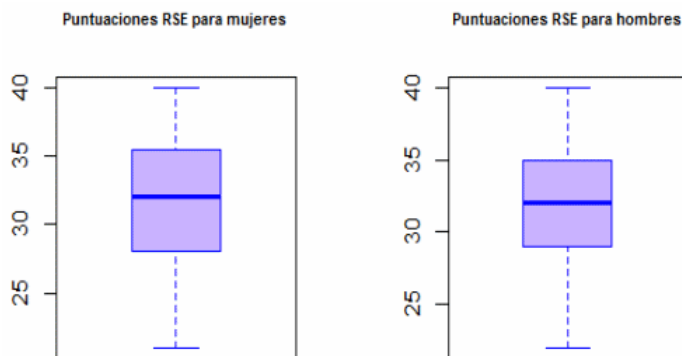


FIGURA 1: Distribución de las puntuaciones de RSE.

Construcción de las listas de palabras

139 personas calificaron la frecuencia de uso y la valencia afectiva de las 60 palabras propuestas en el instrumento. 30 de las 60 palabras fueron elegidas dentro de las que mostraron mayor frecuencia de elección en el Listado de Adjetivos para la Evaluación del Autoconcepto – LAEA (Garaigordobil 2008) y las 30 restantes fueron sus opuestos.

Luego del proceso de recolección de información se obtuvo que de las 60 palabras 27 fueron evaluadas como negativas (puntuaciones comprendidas entre 2.93 y 3.35), 26 fueron evaluadas como positivas (puntuaciones comprendidas entre 5.16 y 6.10) y 7 quedaron clasificadas como neutras (puntuaciones entre 4.07 y 4.49). De los grupos de características positivas y negativas se eligieron las 22 que contaron con los mayores puntajes para frecuencia de uso.

Siguiendo el procedimiento dispuesto por Woolfolk et Al (1995) las palabras negativas se distribuyeron aleatoriamente en 2 sub-listas (A y B) y se realizó la misma acción con las palabras positivas y se crearon dos sub-listas más (C y D). Se procedió a configurar las listas que formarían parte de la tarea. Para la lista 1, que tendría más características negativas se usaron las sub-listas A, B y C. Para la lista 2, conformada por mayor cantidad de palabras positivas se usaron las sub-listas A, C y D y para la lista 3 que debía ser neutra se usaron las sub-listas A, C y 5 palabras provenientes de la sub-lista B y 6 palabras de la sub-lista D elegidas aleatoriamente.

Aplicación de la “Self-Complexity Task”

Ajustándose al procedimiento de Woolfolk et Al (1995) 60 personas eran requeridas para la aplicación de la “Self-Complexity task” (tabla 1). De acuerdo a los resultados del pre-test se eligieron 75, contando con los 15 excedentes para cubrir la posible no respuesta o aumentar el número de elementos por grupo. Las personas elegidas a través del pre-test fueron asignadas aleatoriamente a cada uno de los tres grupos posibles por tipo de lista de palabras, teniendo en cuenta el sexo y el puntaje en obtenido en el RSE. Finalmente la aplicación se realizó con 47 que se distribuyeron en los grupos tal y como se muestra en la tabla 2.

La tarea estuvo disponible en línea y se compartió con los participantes a través de un enlace que les fue enviado al correo electrónico.

Los datos obtenidos fueron procesados con el estadístico H (Ver ecuación 1) y posteriormente se procedió al análisis de varianza. Dado que no se obtuvo información del mínimo de sujetos requerido para cada grupo, se hizo necesario realizar el cálculo de los cuadrados medios del ANOVA marginal o tipo III, útil para casos desbalanceados, disponiendo de tres factores: el tipo de lista (3 niveles: Negativa – Lista 1, Positiva – Lista 2 y Neutra _ Lista 3), el sexo (2 niveles: Femenino y Masculino) y el nivel de autoestima (2 Niveles: Bajo y Alto).

TABLA 1: Configuración de grupos (Woolfolk, Novalany, Gara, Allen y Polino 1995)

	Lista 1		Lista 1		Lista 1	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Nivel Bajo RSE	5	5	5	5	5	5
Nivel Alto RSE	5	5	5	5	5	5
Total	20		20		20	

Fuente: Autor basado en Woolfolk et al. (1995). *Self-complexity, self-evaluation, and depression: An examination of form and content within the self-schema. Journal of Personality and Social Psychology.*

TABLA 2: Configuración de grupos en el presente estudio.

	Lista 1		Lista 1		Lista 1	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
Nivel Bajo RSE	5	4	5	6	3	2
Nivel Alto RSE	5	6	4	3	2	2
Total	20		18		9	

Fuente: Autor.

4. Desarrollo.

La variable respuesta SC presentó puntajes que oscilaron entre los 0.33 y 3.47, con una media aritmética igual a 1.62 y una desviación estándar de 0.62. El número mínimo de aspectos formulados fue 1 y el máximo 7. Cada persona usó en promedio 9.87 de las 33 características posibles para describirse.

En los grupos dados por tipo de lista las medias y desviaciones estándar fueron respectivamente para el grupo 1 (lista negativa) de 1.57 y 0.42, para el grupo 2 (lista positiva) de 1.79 y 1, y para el grupo 3 (lista neutra) de 1.56 y 0.47. Con esta tendencia en los datos y como puede observarse en la gráfica 2, es posible anticipar que no existirá homocedasticidad.

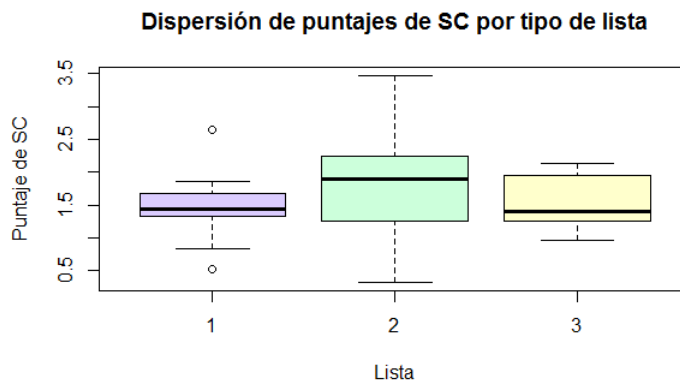


FIGURA 2: Diagrama de cajas por tipo de lista

Análisis de varianza

Para la realización del ANOVA se procedió en primer lugar a verificar los requisitos de normalidad y homocedasticidad. En la confirmación del supuesto de normalidad se realizó una gráfica Q-Q (Gráfica 3) en la que se evidencia un adecuado ajuste de la variable respuesta a la distribución normal.

Respaldo este supuesto se realizó la prueba de Shapiro-Wilk con la cual se corroboró que se puede asumir una distribución normal para SC ($W = 0.967, p - \text{valor} = 0.212$).

En lo referente a la homogeneidad de varianzas se usó la prueba de Levene para la variable respuesta agrupada por el tipo de lista, obteniendo un valor F de 3.46 con un p-valor de 0.03, con lo que se puede considerar que no existe homocedasticidad.

Con el fin de disminuir la diferencia entre las varianzas de los grupos se realizó la transformación de la

variable respuesta usando su raíz cuadrada. Con estos datos, la prueba de Levene evidenció homocedasticidad ($F = 2.671$, $p - \text{valor} = 0.08$) y la normalidad de la distribución se mantuvo (prueba Shapiro-Wilk $W = 0.98$, $p - \text{valor} = 0.612$).

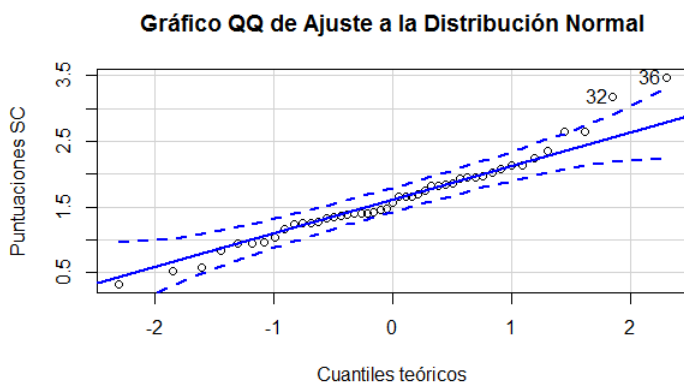


FIGURA 3: Gráfico QQ variable SC.

Una vez culminada la etapa de verificación de supuestos, se realizó el análisis de varianza para grupos desbalanceados y con un 95 % de confianza no se halló evidencia que permita sostener que existen diferencias estadísticamente significativas en las medias para ninguna de las fuentes de variación observadas en el estudio. Los valores obtenidos se ven en la tabla 3.

TABLA 3: ANOVA para los factores tipo de lista, sexo y auto-estima. $\alpha = 0.05$

Fuente de variación	Suma de Cuadrados	Grados de Libertad	F Teórico	F Calculado	P Valor
Lista	0.00045	2	3267	0.0034	0.9966
Sexo	0.00323	1	4121	0.0487	0.8266
Autoestima	0.00009	1	4121	0.0013	0.9710
Lista*Sexo	0.01766	2	3267	0.1331	0.8758
Lista*Autoestima	0.01100	2	3267	0.0829	0.9206
Sexo*Autoestima	0.00639	1	4121	0.0964	0.7581
Lista*Sexo*Autoestima	0.00871	2	3267	0.0657	0.9366
Residual	7.4396	35			

5. Discusión

La construcción de una versión en castellano de la “Self-Complexity Task” requiere de un proceso complejo que empieza con la definición de la línea teórica y metodológica más conveniente. Las dos líneas conceptuales que con mayor frecuencia se usan son la de Linville (1987) y la de Woolfolk et al (1995). En el presente estudio se planteó la réplica de uno de los seis sub-estudios adelantados por Woolfolk et al (1995) para evaluar la validez de la tarea.

En el estudio original adelantado por Woolfolk et al (1995), se halló evidencia para respaldar el efecto que tenía el tipo de lista usada ($F_{2,48} = 7.02$, $p - \text{valor} = 0.02$), el sexo ($F_{1,48} = 7.95$, $p - \text{valor} = 0.008$)

y la auto-estima ($F_{1,48} = 4.81$, p -valor = 0.04) en las puntuaciones de SC. En contraste, en este estudio ninguna de estas tres variables presentó un efecto significativo.

Teniendo en cuenta estos resultados y de cara al proceso de la posterior adaptación del instrumento al castellano, se supondría como lo más recomendable seguir la línea metodológica de Linville (1987) y construir una tarea con 33 características que hayan mostrado alta frecuencia de uso dentro de la población a estudiar, sin que sea relevante la valencia afectiva de estas palabras.

Sin embargo, es necesario tener presente que este trabajo no abordó la totalidad de análisis adelantados por Woolfolk et al (1995), dado que se replicó tan sólo uno de los seis sub-estudios diseñados en su investigación. Del mismo lado, aunque para la presente investigación se procuró realizar las actividades tal y como fueron descritas en el documento original, no fue posible adelantar una aplicación controlada de los instrumentos, como tampoco, obtener la información de todas las unidades requeridas y el cálculo del ANOVA tuvo que hacerse con un procedimiento para grupos desbalanceados, desencadenando esto en diferencias metodológicas en las que podría encontrarse la discrepancia entre los resultados de los estudios.

También, es necesario considerar como causas, las diferencias socio-culturales de las poblaciones abordadas (Pilarska y Suchańska 2015) y el planteamiento estadístico hecho en el estudio original, especialmente lo referido a la toma de decisiones asociada al cumplimiento de los requerimientos para la ejecución de análisis de varianza, información que no fue reportada en el estudio de Woolfolk et al (1995).

Por otra parte, se confirma que SC tiende a ajustarse a la distribución normal, como ya se había visto en estudios previos y que con la cantidad adecuada de unidades de análisis podría controlarse la diferencia entre las varianzas de los grupos, siendo esta una información relevante para la ejecución de futuros trabajos.

Por estos motivos es conveniente continuar con la investigación sobre “Self-Complexity”, en primera medida, acerca de las condiciones necesarias para garantizar la validez y confiabilidad de la tarea, y continuando posteriormente con la revisión de los sistemas de cálculo diferentes a la H estadística y que han sido planteados por diferentes autores como estrategias más apropiadas para la medición del constructo.

Referencias Bibliográficas

- Aparicio, J., Martínez, M. y Morales, J. (2004), ‘Modelos lineales aplicados en r’, *Dto. Estadística, Matemáticas e Informática*.
- Atienza, F., Moreno, Y. y Balaguer, I. (2000), ‘An análisis of the dimensionality of the rosenberg selfesteem scale in a sample of valencian adolescents’, *Universitas Tarraconensis* **22**, 1–2.
- Campbell, J. D., Chew, B. y Scratchley, L. S. (1991), ‘Cognitive and emotional reactions to daily events: the effects of self-esteem and self-complexity’, *Journal of Personality* **59**(3), 473–505.
- de la Garza García, J. (1999), *Análisis de la información mercadológica: a través de la estadística multivariante*, Alhambra Mexicana.
- De la Garza García, J., Morales Serrano, B. y González Cavazos, B. (2013), *Análisis estadístico multivariante.*, McGraw Hill.
- Fankhauser, B. L. (1991), Self-complexity and family caregivers of Alzheimer’s disease victims, PhD thesis, Texas Tech University.
- Faraway, J. J. (2014), *Linear models with R*, CRC press.
- Garaigordobil, M. (2008), ‘Laea’, *Listado de adjetivos para la evaluación del autoconcepto en adolescentes y adultos*.
- Heath, W. H. (2011), ‘Does self-complexity predict dishonest behavior via cognitive dissonance?’.

- Linville, P. W. (1985), 'Self-complexity and affective extremity: Don't put all of your eggs in one cognitive basket', *Social cognition* **3**(1), 94–120.
- Linville, P. W. (1987), 'Self-complexity as a cognitive buffer against stress-related illness and depression.', *Journal of personality and social psychology* **52**(4), 663.
- McConnell, A. R., Strain, L. M., Brown, C. M. y Rydell, R. J. (2009), 'The simple life: On the benefits of low self-complexity', *Personality and Social Psychology Bulletin* **35**(7), 823–835.
- Monleon-Getino, A. (2017), 'Anova caso no balanceado: suma de cuadrados de tipo i, ii, iii y iv'.
- Niedenthal, P. M., Setterlund, M. B. y Wherry, M. B. (1992), 'Possible self-complexity and affective reactions to goal-relevant evaluation.', *Journal of Personality and Social Psychology* **63**(1), 5.
- Pilarska, A. y Suchańska, A. (2015), 'Self-complexity and self-concept differentiation—what have we been measuring for the past 30 years?', *Current Psychology* **34**(4), 723–743.
- Rosenberg, M. (2015), *Society and the adolescent self-image*, Princeton university press.
- Sakaki, M. (2004), 'Effects of self-complexity on mood-incongruent recall', *Japanese Psychological Research* **46**(2), 127–134.
- Sokal, R. R., Rohlf, F. J., Lahoz León, M. et al. (1979), *Biometría: principios y métodos estadísticos en la investigación biológica*.
- Valiente, L. P. y Tejedor, I. H. (2010), *Bioestadística sin dificultades matemáticas*, Ediciones Díaz de Santos.
- Woolfolk, R. L., Novalany, J., Gara, M. A., Allen, L. A. y Polino, M. (1995), 'Self-complexity, self-evaluation, and depression: An examination of form and content within the self-schema.', *Journal of Personality and Social Psychology* **68**(6), 1108.



ANÁLISIS DEL AUSENTISMO LABORAL POR INCAPACIDADES MÉDICAS EN UNA EMPRESA SIDERÚRGICA DE COLOMBIA EN EL AÑO 2017

Especialización en Estadística

YULI STEFANY SANDOVAL VELANDIA^{1,a}

¹ESCUELA DE POSGRADOS, SECCIONAL DUITAMA, UNIVERSIDAD PEDAGÓGICA Y TECNOLÓGICA DE COLOMBIA, DUITAMA, COLOMBIA

Resumen

En este artículo se presenta la descripción univariada y multivariada de las variables que se encuentran en la base de datos consolidada, la cual contiene registros de ausentismo laboral por incapacidad médica de cargos operativos en una empresa siderúrgica de Colombia en el año 2017. Para el desarrollo de este trabajo, se aplica análisis descriptivo de los datos, tablas de contingencia, medidas de asociación y el análisis de correspondencias múltiple. Finalmente se presentan las conclusiones aplicadas al contexto de la información analizada, proponiendo recomendaciones a la organización y la bibliografía en la que se basó la redacción de este documento.

Palabras clave: Ausentismo laboral, descripción univariada, tablas de contingencia, prueba independencia, medidas de asociación, análisis de correspondencias múltiple (ACM).

Abstract

This article presents the univariate and multivariate description of the variables found in the consolidated database, which contains records of labor absenteeism due to medical disability of operative positions in a steel company in Colombia in the year 2017. For the development of this work, descriptive analysis of the data, contingency tables, association measures and multiple correspondence analysis are applied. Finally, the conclusions applied to the context of the analyzed information are presented, proposing recommendations to the organization and the bibliography on which the writing of this document was based (MCA).

Key words: Labor absenteeism, univariate description, contingency table, test of independence, measures of association, multiple correspondence analysis.

1. Introducción

A nivel empresarial, el ausentismo laboral es un asunto de plena actualidad para las organizaciones, por las graves consecuencias que tiene para la economía y las empresas en términos de productividad y competitividad. El ausentismo es una cuestión bastante compleja, con diferentes conceptualizaciones, es multicausal y produce diversidad de consecuencias (Peña Peña 2016).

En cualquiera de los casos en que suele presentarse el ausentismo laboral en una empresa, influye en el proceso de desarrollo institucional y organizacional, por ello la identificación de las causas, la prevalencia con

^aEstudiante de Especialización en Estadística. E-mail: yulisandoval16@hotmail.com

que se presenta dicho fenómeno debe representar un propósito en las organizaciones modernas y amerita ser objeto de estudio de los profesionales del área salud ocupacional y de preocupación necesaria por la gerencia (Sánchez 2015).

Por ejemplo, en Argentina, debido a la relevancia que presenta el ausentismo para los actores del mundo laboral, el Ministerio de Trabajo, Empleo y Seguridad Social (MTE y SS) decidió incorporar esta temática a la Encuesta de Indicadores Laborales (EIL) a partir de agosto de 2012, con el fin de elaborar una medición agregada sobre este fenómeno para el conjunto de empresas privadas (Ministerio de Trabajo Empleo y Seguridad Social 2015).

En Colombia, el (Ministerio del trabajo 2014), según el Decreto número 1443 de 2014 por el cual se dictan disposiciones para la implementación del Sistema de Gestión de la Seguridad y Salud en el Trabajo (SG-SST), expresa en el Artículo 22, que para la definición y construcción de los indicadores que evalúan el resultado del SG-SST, el empleador debe considerar entre otros, el análisis de los registros de enfermedades laborales, incidentes, accidentes de trabajo y ausentismo laboral por enfermedad.

Lo anterior, sumando que los procedimientos que se tienen que llevar a cabo con las EPS en Colombia para el cobro de las incapacidades, la aprobación de autorizaciones para los procedimientos médicos y las consultas con especialistas requieren de largos tramites que aumentan los costos del ausentismo, el presente estudio busca analizar los registros de ausentismo laboral en una Empresa Siderúrgica en el año 2017, pues se encuentra que más del 50 % de los trabajadores del área operativa presentaron por lo menos un registro de ausentismo por incapacidad médica.

A pesar de que la literatura en el tema es amplia, pues en este se interesan autores de áreas como la salud ocupacional, recursos humanos, entre otras, en cuanto estudios y aplicaciones estadísticas sobre ausentismo laboral en industrias siderúrgicas los autores son muy pocos. Lo que dificulta justificar propuestas a planes de acción en industrias de este gremio, que permitan mitigar las consecuencias del ausentismo por incapacidad y mejorar la calidad de vida de los trabajadores que las presentan.

En consecuencia, se busca dar respuesta al siguiente interrogante: ¿Cuál es la relación entre las incapacidades registradas en una empresa Siderúrgica de Colombia en el año 2017 y las características de los colaboradores que las presentan? Para el desarrollo de esta propuesta se plantea un análisis descriptivo univariado y multivariado de las variables a las que fue posible tener acceso. Se presenta en primer lugar, el análisis descriptivo univariado para las variables que brindan las incapacidades registradas y las variables que caracterizan a quienes las registran. En segundo lugar, la aplicación de tablas de contingencia que buscan asociación entre ausentismo laboral dado en días y las demás variables que se tienen. Por último, se exponen los resultados obtenidos en el análisis de correspondencias múltiple (ACM) de las variables que luego de los resultados de las tablas de contingencias fueron incluidas en esta etapa.

2. Referente Conceptual

2.1. Definiciones del ausentismo laboral

Según (Chiavenato, Mascaró Sacristán, Roa et al. 2007), el absentismo (también llamado ausentismo) es una expresión empleada para designar las faltas o ausencias de los empleados en el trabajo. En un sentido más amplio, es la suma de los periodos en que los empleados se encuentran ausentes en el trabajo ya sea por falta, retardo o algún otro motivo. Argote (como se citó en (Sánchez 2015)), define el ausentismo como los momentos en que el trabajador permanece ausente cuando debiera estar cumpliendo sus compromisos laborales, o bien llamado ausentismo o por otros absentismo laboral, contempla los permisos, incapacidades y retrasos. Espinosa (como se citó en (Sánchez 2015)), complementa afirmando que no se incluyen las ausencias derivadas de las vacaciones, las licencias de maternidad o paternidad, o las licencias por comisión de estudio.

2.2. Clasificación del ausentismo laboral

La Norma Técnica Colombiana (INCONTEC 1996), que tiene por objeto conceptualizar y establecer lineamientos para la clasificación, el registro y el seguimiento estadístico del ausentismo en las empresas del país, presenta la siguiente clasificación:

1. Causas no relacionadas con la salud:
 - Legales pactadas en el Código Sustantivo del Trabajo.
 - Extralegales pactadas en convenciones y acuerdos extralegales.
 - Otras formas de ausentismo autorizadas o no y de cualquier duración. Detenciones, problemas interpersonales, problemas familiares, etc.
2. Causas relacionadas con la salud:
 - Accidente Común: Evento de salud repentino no relacionado con el trabajo, incluye sucesos en el hogar, vía pública, recreativos, entre otros.
 - Accidente de Trabajo: Todo suceso repentino que sobrevenga por causa o con ocasión del trabajo, y que produzca en el trabajador lesión orgánica, perturbación funcional, invalidez o muerte.
 - Enfermedad General: Todo estado patológico que no esté considerado como enfermedad profesional.
 - Enfermedad Profesional: Todo estado patológico permanente o temporal que sobrevenga como consecuencia obligada y directa de la clase de trabajo que desempeña el trabajador o del medio en que se ha visto obligado a trabajar.
 - Control Preventivo.

Para el desarrollo de este estudio, se toma en consideración la segunda clasificación del INCONTEC con respecto a causas relacionadas con la salud, sin tomar en cuenta los controles preventivos. Pues este estudio se baja en los registros de incapacidades.

2.3. Análisis descriptivo univariado

Para el análisis univariado el cual se encarga de las variables por separado, es importante tener presente la escala de medición pues esta indica las medidas que se pueden obtener de la variable dependiendo su naturaleza.

Como afirma Coronado (2013):

Las escalas o niveles de medición se utilizan para medir variables o atributos. Por lo general, se distinguen cuatro escalas o niveles de medición: nominal, ordinal, intervalos y escalas de proporción, cociente o razón. Las dos primeras (nominal y ordinal) se conocen como escalas categóricas, y las dos últimas (intervalo y razón) como escalas numéricas. Las escalas categóricas se usan comúnmente para variables cualitativas, mientras que las numéricas son adecuadas para la medición de variables cuantitativas (Coronado 2013, pág. 106).

2.4. Tablas de contingencia

Los conceptos de tablas de contingencia, pruebas de independencia y medidas de asociación son tomados del Capítulo 2 de Díaz y Morales (2002) Una tabla de contingencia de es un arreglo bidimensional, de una variable fila con f -categorías o modalidades y una variable columna con c -categorías, donde hay fxc celdas, las entradas de las celdas son las frecuencias o conteos del número de casos en cada una de las combinaciones de valores de ambas variables. En general, se nota con n_{ij} a la frecuencia de la i -ésima modalidad de la variable fila y j -ésima de la variable columna. El total por fila o por columna está formado por las frecuencias marginales, y se notan por $n_{i\cdot}$ (donde el punto señala que se suman columnas dentro de la fila i) y $n_{\cdot j}$ (donde el punto señala que se suman filas dentro de la columna j), respectivamente. La suma de las frecuencias por celda es igual a la suma de las frecuencias marginales e igual al número total de individuos seleccionados y

clasificados; se nota por n . La notación general, para una tabla de contingencia de f -filas y c -columnas, se muestra en la figura 1.

Filas	Columnas						Total($n_{i.}$)
	1	2	...	j	...	c	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2c}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ic}	$n_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
f	n_{f1}	n_{f2}	...	n_{fj}	...	n_{fc}	$n_{f.}$
Total($n_{.j}$)	$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.c}$	$n_{..} = n$

FIGURA 1: Tabla de contingencia. Fuente: (Díaz & Morales, 2002)

Donde:

- La frecuencia de la i -ésima modalidad de la variable fila y la modalidad j -ésima de la variable columna se escribe como x_{ij} .
- El total de observaciones en la i -ésima modalidad de la variable fila se nota por $n_{i.}$, es decir,

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{ic} = \sum_{j=1}^c n_{ij} \tag{1}$$

- El total de observaciones en la j -ésima modalidad de la variable columna se nota por $n_{.j}$; es decir,

$$n_{.j} = n_{1j} + n_{2j} + \dots + n_{fj} = \sum_{i=1}^f n_{ij} \tag{2}$$

- El número total de observaciones en la muestra se escribe con n , y es igual a la suma de los márgenes fila o columna, es decir,

$$n = \sum_{i=1}^f \sum_{j=1}^c n_{ij} \tag{3}$$

Por otra parte, las frecuencias pueden ser transformadas en proporciones o porcentajes. Un primer porcentaje se obtiene de dividir cada frecuencia n_{ij} por el número total de observaciones n ; este porcentaje se escriben como f_{ij} , es decir,

$$f_{ij} = \frac{n_{ij}}{N} \times 100 \tag{4}$$

la cantidad f_{ij} corresponde a la proporción o porcentaje de elementos que tienen los atributos i y j .

El segundo porcentaje se obtiene al dividir cada frecuencia n_{ij} por la respectiva frecuencia marginal fila $n_{i.}$, así:

$$f_{i|j} = \frac{n_{ij}}{n_{i.}} \times 100 \tag{5}$$

La cantidad $f_{j|i}$ es la proporción de elementos de cada celda, respecto al total de la fila i estas frecuencias corresponden al perfil fila. El tercer porcentaje se obtiene al dividir cada frecuencia n_{ij} por la respectiva frecuencia marginal columna $n_{.j}$:

$$f_{ij} = \frac{n_{ij}}{n_{.j}} \times 100 \tag{6}$$

La cantidad $f_{i|j}$ es la proporción de elementos de cada celda, respecto al total de la columna j estas frecuencias corresponden al perfil columna.

2.4.1. Pruebas de independencia

Una vez que se han dispuesto los datos en una tabla de contingencia, la pregunta más importante, en general, es si las variables cualitativas que conforman la tabla de contingencia son independientes o no.

La hipótesis nula de independencia está dada por:

$$H_0 = \text{La variable fila es independiente de la variable columna} \quad (7)$$

La estadística de prueba que es empleada en el juzgamiento de esta hipótesis es:

$$X^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (8)$$

que bajo la hipótesis nula de independencia, tiene distribución de probabilidad ji -cuadrado con $(f-1) \times (c-1)$ grados de libertad.

Se rechaza la hipótesis nula a un nivel α cuando se verifica que $X_0^2 > X_{(f-1)(c-1),\alpha}^2$

2.4.2. Medidas de asociación

A continuación, algunas medidas relacionadas con la estadística ji -cuadrado, como se puede ver en Díaz y Morales (2002) algunas de estas son:

El coeficiente de contingencia

Es una medida del grado de asociación o relación entre dos conjuntos de atributos. Es especialmente útil cuando se tiene información clasificadora acerca de uno o ambos conjuntos de atributos. El grado de asociación entre dos conjuntos de atributos, sean ordinales o no, se puede describir mediante la siguiente fórmula:

$$C = \sqrt{\frac{X_0^2}{X_0^2 + n}}, \text{ donde } X^2 = \sum_{i=1}^f \sum_{j=1}^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \quad (9)$$

La estadística C toma valores entre 0 y 1. Valores cercanos a 0 muestran una baja asociación entre las variables, mientras que valores próximos a 1 indican una posible alta asociación.

El coeficiente (V) de Cramer

Este coeficiente tiene un valor máximo en tablas de contingencia de cualquier tamaño. Se define como:

$$V = \sqrt{\frac{X_0^2}{nk}} \quad (10)$$

donde $k = mnf - 1$, $c - 1$ es el menor número de modalidades fila (o columna) menos uno de la tabla de contingencia. Se trata de un coeficiente que toma el valor 1 cuando hay asociación perfecta entre los atributos, cualquiera que sea el tamaño de la tabla de contingencia.

Análisis de correspondencias múltiple (ACM)

El análisis de correspondencias es de aplicación con dos caracteres o variables cualitativas, cada una de las cuales puede presentar varias modalidades o categorías. Pero el método es generalizable al caso de un número de variables o caracteres cualitativos mayor de dos en el análisis de correspondencias múltiple, el cual permite estudiar las relaciones entre las modalidades de todas las características cualitativas consideradas (Pérez 2004). Lo correspondiente a referente teórico del ACM se continúa tomando de Pérez (2004). En el análisis de correspondencias múltiples se ordenan los datos en una tabla Z denominada tabla disyuntiva completa (ver Figura 2) que consta de un conjunto de individuos $I = 1, \dots, i, \dots, n$ (en filas), un conjunto de

variables o caracteres cualitativos $J_1, \dots, J_k, \dots, J_Q$ (en columnas) y un conjunto de modalidades excluyentes $1, \dots, m_k$ para cada carácter cualitativo. El número total de modalidades será entonces:

$$J = \sum_{k=1}^Q m_k \quad (11)$$

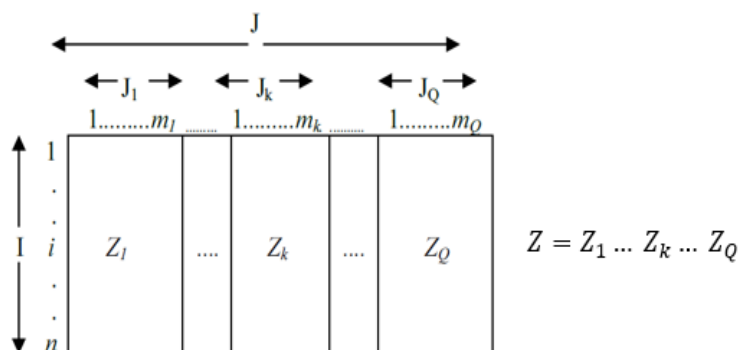


FIGURA 2: Tabla disyuntiva completa Z. Fuente: (Pérez, 2004)

El elemento Z_{ij} de la tabla toma el valor 0 o 1 según que el individuo i haya elegido (esté afectado por) la modalidad j o no. Por lo tanto, cada rectángulo de la tabla disyuntiva completa puede considerarse, aunque no lo sea, como una tabla de contingencia cuyos elementos son 0 o 1. La tabla disyuntiva completa Z consta entonces de Q subtablas yuxtapuestas, con la finalidad de obtener una representación simultánea de todas las modalidades (columnas) de todos los individuos (filas).

$$Z_{ij} = k_{ij} = 0 \text{ ó } 1 \quad (12)$$

$$k_{i.} = \sum_j k_{ij} = Q = \text{número de modalidades} \quad (13)$$

$$k_{.j} = \sum_i k_{ij} = \text{número de individuos que poseen la modalidad } j \quad (14)$$

$$\frac{f_{ij}}{f_{i.}} = \frac{k_{ij}}{k_{i.}} = \frac{1}{Q} = \text{inverso del número de modalidades} \quad (15)$$

Para obtener los factores es necesario diagonalizar la matriz

$$V = \frac{D - 1B}{Q} \quad (16)$$

$$B = Z'Z = \text{Matriz simétrica formada por } Q^2 \text{ bloques} = \text{Tabla de Burt.} \quad (17)$$

D = Matriz diagonal cuyos elementos diagonales son los de la matriz de Burt, siendo nulos el resto de los elementos.

Las fórmulas de transición que permiten representar simultáneamente los puntos línea y los puntos columna sobre los mismos gráficos relacionando así los resultados en los dos subespacios tomarán ahora las siguientes expresiones:

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{Q} \sum_{j=1}^p k_{ij} G_\alpha(j) \quad (18)$$

$$G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{k_{.j}} \sum_{i=1}^n k_{ij} F_\alpha(i) \quad (19)$$

Donde

$$F_{\alpha}(i) = \text{La proyección de un punto individuo } i \text{ sobre el eje } \alpha. \quad (20)$$

$$G_{\alpha}(j) = \text{La proyección de un punto modalidad } j \text{ sobre el eje } \alpha. \quad (21)$$

$$\frac{1}{\sqrt{\lambda_{\alpha}}} = \text{Coeficiente de dilatación.} \quad (22)$$

Como el Análisis Factorial de Correspondencias es centrado y el centro de gravedad de las modalidades de una variable ($\frac{1}{\sqrt{n}}$) coincide con el de la nube de puntos variables, y con el origen, las modalidades de cada variable están centradas en torno al origen, no pudiendo tener todo el mismo signo. La parte de inercia debida a una modalidad j es mayor cuanto menor sea el efectivo de esa modalidad. Si G representa el centro de gravedad, la inercia debida a la modalidad j viene dada por:

$$I(j) = \frac{1}{Q} \left(1 - \frac{k \cdot j}{n}\right) \quad (23)$$

Por lo tanto, es aconsejable eliminar las modalidades elegidas muy pocas veces, construyendo otra modalidad uniéndola a la más próxima. La parte de inercia debida a una variable es función creciente del número de modalidades de respuesta que tiene, ya que la inercia de una variable es la suma de las inercias de sus modalidades:

$$I(J_k) = \sum_{j \in J_k}^n I(j) = \frac{1}{Q} (m_k - 1) \quad (24)$$

Si una variable tiene un número de modalidades demasiado grande, conviene reagrupar las, modalidades en un número que sea razonable y mantener el sentido, para evitar así influencias extremas.

La inercia total es la suma de las inercias de todas las modalidades:

$$I = \sum_{j \in J_k} I(J_k) = \frac{J}{Q} - 1, \quad (25)$$

donde $\frac{J}{Q}$ es el número medio de modalidades por variable cualitativa o carácter.

3. Metodología

La aplicación se desarrolló con enfoque cuantitativo y tipo de investigación exploratoria, basada en la consolidación de dos bases de datos, con uso autorizado por parte de la empresa manteniendo su confidencialidad. La primera base de datos presenta los registros de las incapacidades de 2017, donde se encuentran variables como: días de incapacidad, clase de incapacidad, tipo de incapacidad, fecha de la incapacidad y diagnóstico. La segunda base de datos presenta información sobre todos los trabajadores de la organización, donde se encuentran variables como: edad, antigüedad, área y cargo. Se eliminaron otras variables que las bases de datos contenían y que no eran pertinentes para el estudio, dejando únicamente las mencionadas anteriormente. La variable sexo no fue incluida pues todos los registros corresponden al sexo masculino.

Una vez consolidada la base de datos, de los 534 registros que presenta se descartan 81 por inconsistencias, lo que corresponde a una muestra del 84.83 %. Esta muestra contiene los registros de incapacidad y datos de los trabajadores del área operativa que presentaron al menos una incapacidad en una empresa siderúrgica de Colombia en el año 2017.

Se identifica la escala de medición a la que pertenece cada una de las variables para aplicar el análisis univariado. Con base en la descripción, las variables codifican con el fin de reducir el número de categorías (como en el caso de la variable diagnóstico) y cambiar de escala a las variables cuantitativas. Una vez que todas variables se encuentran en escala nominal u ordinal, se procede con la aplicación de las tablas de contingencia, pruebas de independencia, medidas de asociación.

Luego de obtener los resultados de asociación, se realiza el análisis de correspondencias, pues como se afirma en Carvajal, Carpintero y Mejía (2006), a través de la prueba de asociación o independencia de la X^2 (test de chi-cuadrado) se extrae información sobre la relación significativa o no entre un par de variables,

sin aclarar cuáles categorías de ellas están implicadas; mientras en el análisis de correspondencias se logra establecer relaciones entre categorías y definir similitudes o disimilitudes entre ellas, lo que permite su agrupamiento si se detecta que se corresponden y todo queda plasmado en un espacio dimensional de escasas variables sintéticas o factores que pueden ser interpretados y que además condensan al máximo la información, la cual suele ser proporcionada en modo gráfico (valores asociados próximos).

Sin embargo, es necesario realizar las pruebas de asociación entre las variables ya que como afirman Díaz y Garrido (2015), si no hay relación entre las variables no tiene sentido realizar un análisis de correspondencias.

Todos los cálculos que se presentan en esta aplicación se desarrollaron con el software libre R.

4. Resultados

4.1. Descripción de variables

Usando la función summary de R, se obtiene la Figura 3 que resume las variables dependiendo de su escala de medición

Días		Edad		Antigüedad		Fecha		Clase		Tipo		Diagnostico		Area		Cargo	
Min. : 1.000	Min. : 23.00	Min. : 0.00	mayo-mayo	: 48	Inicial :345	AccidentedeTrabajo	: 25	M545	: 39	1:195	43	: 86					
1st Qu.: 2.000	1st Qu.:34.00	1st Qu.:10.00	diciembre-diciembre:	47	Prorroga:108	EnfermedadGeneral	:412	A09X	: 21	2:160	23	: 53					
Median : 3.000	Median :38.00	Median :11.00	julio-julio	: 35		EnfermedadProfesional:	14	M332	: 13	3: 82	40	: 53					
Mean : 6.289	Mean :40.12	Mean :12.92	junio-junio	: 35		SOAT	: 2	M751	: 13	4: 16	4	: 23					
3rd Qu.: 7.000	3rd Qu.:46.00	3rd Qu.:13.00	octubre-octubre	: 34				G560	: 10		19	: 14					
Max. :30.000	Max. :60.00	Max. :43.00	marzo-marzo	: 31				(Other):	351		(Other):	221					
			(Other)	:223				NA's	: 6		NA's	: 3					

FIGURA 3: Resumen de variables. Fuente: Autora, 2018

A continuación, se contextualiza a cada variable usando como soporte los resultados expuestos en la Figura 3 y como se mencionó en la metodología, se recodifican algunas de las variables (Figura 4).

Variable días: representa los días totales que el médico a quien el trabajador presentó su dolencia le asignó, es decir, representa la duración del ausentismo por incapacidad médica. Como se observa en Figura 3, el tiempo mínimo de una incapacidad es de 1 día y su duración máxima es de 30 días. Presenta varianza de 58.44, desviación estandar 7.64, coeficiente de asimetría 1.94 y curtosis 5.92. Las incapacidades que más se registran son aquellas en las que su duración es de 1,2 o 3 días. Por medio de la frecuencia relativa acumulada se tiene corresponden al 63.53% de las incapacidades registradas y tan solo el 9.2% corresponden a incapacidades de más de 20 días.

Para el cambio de escala de medición, a esta variable se le asignaron 4 categorías: incapacidades con duración de 1 a 2 días, de 3 a 5 días, de 6 a 15 días y de 16 a 30 días.

Variable edad: representa la edad de los trabajadores que registran incapacidad. El más joven de ellos con 23 años y el mayor con 60 años, pero el intervalo de edad que más se registra es de 34 a 46 años. Presenta varianza de 58.96, desviación estandar 7.67, coeficiente de asimetría 0.39 y curtosis 2.53. Las categorías asignadas son: edades de 20 a 30 años, de 31 a 40, de 41 a 50 y de 51 a 60.

Variable antigüedad: representa en años la antigüedad de los trabajadores en la empresa. Presenta varianza de 47.15, desviación estandar 6.86, coeficiente de asimetría 1.63 y curtosis 5.95. Aparece que existe antigüedad de 0, pero esto quiere decir que el trabajador no tiene más de un año de vinculación con la empresa. Los trabajadores llevan en promedio 12.9 años vinculados a la empresa. Las categorías asignadas son: de 0 a 5, de 6 a 10, de 11 a 15, de 16 a 20, de 21 a 25, de 26 a 30 y de 31 a 43 años.

Variable fecha: Todas las categorías de esta variable no se tomaron en el summary debido a la cantidad, al igual que ocurre en las variables cargo y diagnóstico. Esta variable representa el mes de inicio y fin de la incapacidad, debido a que se pueden presentar incapacidades que tengan fecha de inicio en un mes y finalicen en otro. Se categoriza por los trimestres del año.

Variable clase: Como en la variable días, es el médico quien decide si el trabajador requiere de más incapacidad a lo que se le denomina como prórroga de incapacidad. En caso de que la incapacidad no se prórroga, se denomina inicial.

Variable tipo: La descripción de sus categorías fue presentada en los referentes teóricos, en clasificación del ausentismo. Como se ven en la Figura 3, el 90.99% de los registros corresponden a enfermedad general. Es importante aclarar que los accidentes de trabajo que aparecen no se presentaron en 2017, son incapacidades otorgadas a causa de accidentes de años anteriores o de eventos deportivos de la empresa.

Variable diagnóstico: Se encontró que esta variable presenta 196 categorías diferentes, por lo que se decide recodificarla según la Clasificación internacional de enfermedades 10° (CIE10°) dada por la Organización Mundial de la Salud [OMS]. Encontrando la presencia de 16 capítulos, denominando y usando para posteriores cálculos una nueva variable llamada Capítulo de diagnóstico. Se reagrupan los 9 capítulos en los que su frecuencia absoluta es menor a 7, denominando esta categoría como OTRO para un total de 9 categorías.

- Categorías variable capítulo:

- ◊ I: Ciertas enfermedades infecciosas y parasitarias.
- ◊ VI: Enfermedades del sistema nervioso.
- ◊ X: Enfermedades del sistema respiratorio.
- ◊ XI: Enfermedades del sistema digestivo.
- ◊ XII: Enfermedades de la piel y el tejido subcutáneo.
- ◊ XIII: Enfermedades del sistema osteomuscular y del tejido conjuntivo.
- ◊ XVIII: Síntomas, signos y hallazgos anormales clínicos y de laboratorio no clasificados en otra parte.
- ◊ XIX: Traumatismos, envenenamientos y algunas otras consecuencias de causa externa.
- ◊ OTROS: incluye los capítulos II, IV, VII, VIII, IX, XIV, XVI, XX y XXI.

Variable área: Se incluyen 4 categorías correspondientes al área de gestión a la que pertenecen los trabajadores, estas áreas se continúan denominando del 1 al 4 por tema de confidencialidad.

Variable Cargo: Se presentan 48 de 97 cargos operativos de la empresa. Es importante tener presente esta variable, pero para cálculos posteriores debido al número de categorías será usada únicamente la variable área de gestión, pues están ligadas.

Días	Edad	Antigüedad	Fecha	Capítulo
1:191	20_30: 26	0_5 : 44	1rTrimestre : 97	XIII :144
2: 46	31_40:232	11_15:174	2doTrimestre:140	XIX :102
3:139	41_50:144	16_20: 19	3rTrimestre :101	X : 49
4: 77	51_60: 51	21_25: 17	4toTrimestre:115	OTRO : 38
		26_30: 28		I : 33
		31_43: 12		(other): 81
		6_10 :159		NA's : 6

FIGURA 4: Variables recodificadas. Fuente: Autora, 2018

4.2. Cruce de variables

La tabla 1 presenta los resultados obtenidos del cruce entre la variable días (como se indicó en su descripción representa el ausentismo en días) y las demás variables incluidas en el estudio.

Sobre los valores de la prueba de independencia y los coeficientes de asociación, se observa que las variables edad, clase, capítulo y área de gestión presentan fuerte asociación con los días de incapacidad.

VARIABLES CRUZADAS		P-VALOR	COEFICIENTE DE CONTINGENCIA	COEFICIENTE (V) DE CRAMER
Días	Edad	9.207x10-06	0.283	0.171
	Antigüedad	0.04197	0.247	0.147
	Fecha	0.0882	0.18	0.105
	Clase	2.2x10-16	0.486	0.556
	Capítulo	8.121x10-14	0.445	0.287
	Área de gestión	0.00182	0.234	0.139

TABLA 1: Cruce de variables. Fuente: Autora, 2018

4.3. Análisis de correspondencias múltiple (ACM)

Como se puede ver en la figura 5 de autovalores y porcentaje de varianza total explicada para cada dimensión, con un total de 18 dimensiones, la dimensión 1 es la más importante pues su autovalor es de 0.409521, seguida de la dimensión 2 con 0.273311. Estos autovalores representan la cantidad de variabilidad total que está siendo explicada por las dimensiones, que en total corresponde al 18.9%.

Principal inertias (eigenvalues):

dim	value	%	cum%	scree plot
1	0.409521	11.4	11.4	***
2	0.273311	7.6	18.9	**
3	0.263758	7.3	26.3	**
4	0.250974	7.0	33.2	**
5	0.231019	6.4	39.6	**
6	0.218962	6.1	45.7	**
7	0.212533	5.9	51.6	*
8	0.205786	5.7	57.3	*
9	0.195414	5.4	62.8	*
10	0.188498	5.2	68.0	*
11	0.185432	5.1	73.1	*
12	0.171920	4.8	77.9	*
13	0.169325	4.7	82.6	*
14	0.157192	4.4	87.0	*
15	0.146733	4.1	91.0	*
16	0.131403	3.6	94.7	*
17	0.114891	3.2	97.9	*
18	0.076759	2.1	100.0	*

Total:		3.603429	100.0	

FIGURA 5: Autovalores. Fuente: Autora, 2018.

Debido a que las dimensiones 2 y 3 tienen valores propios cercanos, más adelante se comparan también las dimensiones 1 y 3.

La figura 6 muestra las categorías de las variables seleccionadas para el ACM junto con las coordenadas y los cuadrados de los cosenos en las dos primeras dimensiones. Cuando los cosenos cuadrados toman un valor cercano a 1, indican una alta asociación entre la modalidad y el eje (Díaz y Morales 2002).

	name	k=1	k=2	cor	cor.1
1	Cap:I	-0.898	0.032	0.064	0.000
2	Cap:OTRO	-0.357	0.012	0.012	0.000
3	Cap:VI	-0.831	0.427	0.032	0.008
4	Cap:X	-1.018	0.180	0.127	0.004
5	Cap:XII	-0.252	-1.249	0.004	0.107
6	Cap:XIII	-0.605	-1.302	0.009	0.042
7	Cap:XVIII	0.332	0.081	0.052	0.003
8	Cap:XIX	0.929	0.390	0.253	0.045
9	Cap:XVIII	-0.724	-1.021	0.026	0.051
10	Ed:20_30	-0.052	1.309	0.000	0.105
11	Ed:31_40	-0.238	0.392	0.059	0.160
12	Ed:41_50	0.589	-0.504	0.163	0.119
13	Ed:51_60	-0.576	-1.014	0.042	0.130
14	Ges:Ges1	-0.084	0.523	0.005	0.206
15	Ges:Ges2	0.293	-0.953	0.047	0.499
16	Ges:Ges3	-0.455	0.462	0.045	0.047
17	Ges:Ges4	0.352	0.847	0.005	0.026
18	Dias:1_2	-0.688	-0.230	0.341	0.038
19	Dias:16_a_30	1.632	0.742	0.303	0.063
20	Dias:3_4_5	-0.255	0.471	0.029	0.099
21	Dias:6_a_15	1.170	-0.721	0.282	0.107
22	Cla:Inicial	-0.432	-0.065	0.591	0.014
23	Cla:Prorroga	1.366	0.212	0.588	0.014

FIGURA 6: Coordenadas y cosenos cuadrados. Fuente: Autora, 2018.

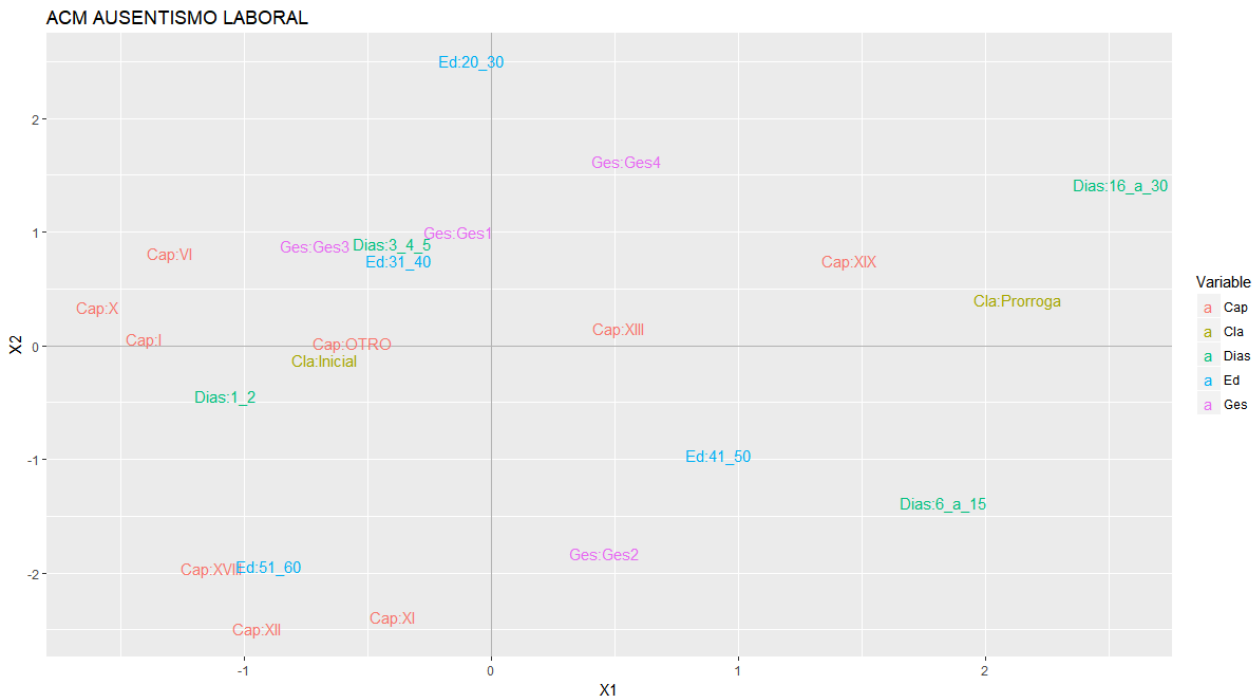


FIGURA 7: Dimensiones 1 y 2 de ACM. Fuente: Autora, 2018.

La primera dimensión se define por la variable Clase de incapacidad, sus coordenadas indican que esta variable diferencia a los individuos entre quienes presentan incapacidad inicial y prórroga de la incapacidad. Observando las coordenadas más cercanas a estas dos categorías de la variable clase de incapacidad, se tiene que las incapacidades iniciales tienen gran afinidad con el área de gestión 3 y las incapacidades que son prórrogas con las categorías de 6 a 15 y 16 a 30 de la variable días de incapacidad.

Para realizar una clasificación más detallada de los individuos, se analiza por cuadrantes la figura 7 obtenida de los resultados de la figura 6.

Cuadrante I: En la figura 8 se evidencia que en los cuadrantes I y IV que corresponden a las incapacidades prórroga no presentan tantos individuos y las coordenadas de las variables no se encuentran tan próximas como en los cuadrantes II y III. Sin embargo, en este cuadrante se posicionan las incapacidades de 16 a 30 días, con el capítulo de diagnóstico XIX y también el área de gestión 4.

Cuadrante II: Se ubican a los individuos que presentan incapacidad inicial, de 3 a 5 días, entre las edades 31 a 40, por los capítulos de diagnósticos I, VI, X y OTRO, quienes en su mayoría pertenecen al área de gestión 1 y 3.

Cuadrante III: Al igual que en el cuadrante II y como se explicó anteriormente, en este cuadrante se ubican a los individuos que presentan incapacidad inicial. Pero a diferencia del anterior cuadrante, se ubican los individuos que presentan incapacidades de 1 a dos días, entre las edades 51 a 60, por los capítulos de diagnósticos XI, XII, y XVIII.

Cuadrante IV: Se ubican a los individuos que presentan prórroga de incapacidad, de 6 a 15 días, entre las edades 41 a 50, quienes en su mayoría pertenecen al área de gestión 2.

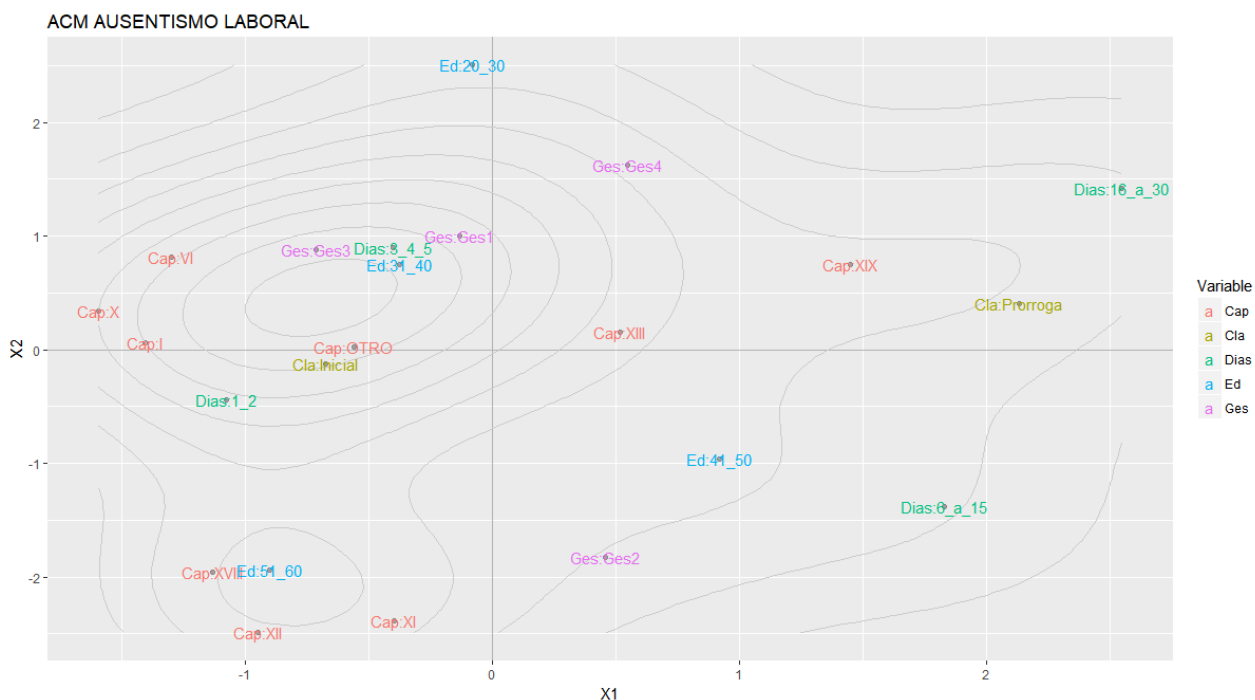


FIGURA 8: Dimensiones 1 y 2 de ACM con individuos. Fuente: Autora, 2018.

Como se expuso al comienzo del desarrollo de ACM, ahora se observan gráficamente las dimensiones 1 y 3 (Figura 9), pues esta combinación explica el 18,7% de la varianza total de los individuos. Se observa, que las categorías de las variables tienen una distribución en los cuadrantes del gráfico similar al gráfico de las dimensiones 1 y 2. Pero en esta oportunidad las coordenadas de las dimensiones 1 y 3 para las categorías área de gestión 3, el rango de edad entre 51 y 60, los capítulos de diagnósticos categorizados como XII y con incapacidad de 3 a 4 días presentan gran asociación. Cabe resaltar que estas categorías se encuentran

ubicadas en el cuadrante III en el que la clase de incapacidad es inicial. Comparando los resultados de los gráficos 2 y 3 se resalta la influencia de las categorías de la variable clase de incapacidad. Se confirma que el gráfico 2 brinda suficiente información, por lo que no se ve la necesidad de tener en cuenta otras dimensiones para conclusiones de la presente aplicación.

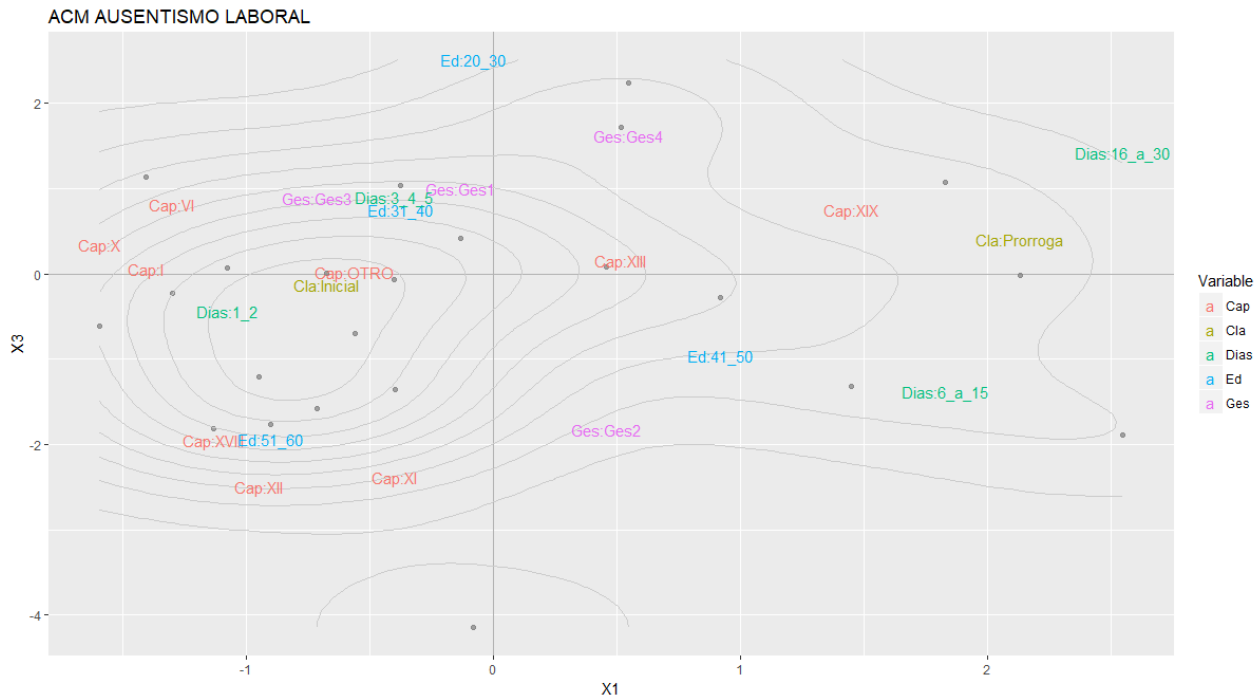


FIGURA 9: Dimensiones 1 y 3 de ACM con individuos. Fuente: Autora, 2018.

5. Conclusiones

Dado a que de los registros que se presentan el 76.15 % son iniciales, el 42.16 % tienen duración de 1 a dos días y a que en Colombia la empresa está en la obligación de cubrir los dos primeros días de la incapacidad cuando esta es inicial, se deduce que los costos que genera el ausentismo laboral por incapacidades en la organización son elevados y amerita reforzar acciones al respecto.

Al describir los resultados del ACM por cuadrantes, es posible diferenciar a los trabajadores en 3 grupos gracias a su presencia en los respectivos cuadrantes y a las relaciones que se generaron entre categorías. El primer grupo se describe en el cuadrante II, el segundo en el cuadrante III y el para el tercer grupo se consolidan los cuadrantes I y IV.

Basado en la agrupación descrita, se recomienda a la organización implementar planes de acción que permitan mitigar el impacto del ausentismo laboral por incapacidades dependiendo de las necesidades de cada grupo. Aunque el objeto de este estudio no es determinar que diagnósticos o grupos de trabajadores son más importantes que otros, se sugiere a la organización que los planes de acción comiencen tratando al grupo 3, pues según el rango de edad de este grupo se estaría actuando sobre el 51.21 % de los trabajadores que presentaron por lo menos 1 registro de ausentismo y al área que presenta condiciones de trabajo más fuertes de las otras.

Por último, se recomienda a la empresa realizar un estudio sobre clima organizacional ya que las áreas que acá se presentaron se distinguen también por los administrativos a cargo y como se demostró en los resultados de la tabla 1 esta variable presenta fuerte asociación con los días de ausentismo.

Referencias Bibliográficas

- Carvajal, P. O., Carpintero, A. A. T. y Mejía, J. (2006), 'Deserción estudiantil, facultad de ciencias de la salud 2000-2004. análisis de correspondencias múltiples', *Revista Médica de Risaralda* **12**(2), 3.
- Chiavenato, I., Mascaro Sacristan, P., Roa, H. et al. (2007), *Administración de recursos humanos: El capital humano de las organizaciones.*, McGraw-Hill Interamericana.
- Coronado, P. J. (2013), 'Escalas de medición', *Paradigmas* **2**(2), 104–125.
- Díaz, I. y Garrido, I. (2015), 'Guía Análisis de Correspondencias Múltiples en SPSS', *Guía Análisis de Correspondencias Múltiples en SPSS* .
- Díaz, L. G. y Morales, M. (2002), 'Análisis estadístico de datos categóricos', *Notas de Clase del Departamento de Estadística de la Universidad Nacional de Colombia. Bogotá: Universidad Nacional de Colombia* .
- INCONTEC (1996), 'Norma Técnica Colombiana NTC 3793', *Norma Técnica Colombiana NTC 3793* .
- Ministerio del trabajo, M. (2014), 'Decreto Número 1443', *Ministerio del trabajo, MINTRABAJO* .
- Ministerio de trabajo Empleo y Seguridad Social, M. (2015), 'Incidencia del ausentismo laboral en las empresas privadas de los principales centros urbanos 2 semestre de 2015', *Incidencia del ausentismo laboral en las empresas privadas de los principales centros urbanos 2 semestre de 2015* .
- Panamericana de la Salud, O. (1995), *Clasificación estadística internacional de enfermedades y problemas relacionados con la salud*, Vol. 1, Pan American Health Org.
- Peña Peña, L. K. (2016), 'Factores y efectos del ausentismo laboral en La Empresa Acerías Paz del Rio SA: propuesta de mejoramiento', *Universidad Pedagógica y Tecnológica de Colombia* .
- Pérez, C. L. (2004), *Técnicas de análisis multivariante de datos*, Pearson Educación.
- Sánchez, D. C. (2015), 'Ausentismo laboral: una visión desde la gestión de la seguridad y la salud en el trabajo', *Rev. salud bosque* **5**(1), 43–54.