



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Identificación de los Factores Principales Asociados con la Elección de un Grado Universitario en el Sistema Educativo Español: Análisis de Componentes Principales y Factorial

Trabajo Fin de Máster

Máster Universitario en Gestión de la Información

Autor: Regina Sandra Mollapaza Quispe

Tutores: Lenin Guillermo Lemus Zúñiga

Eva Onaindia de La Rivaherrera

Curso: 2016-2017

A mis padres por ser los principales
promotores de mis sueños.

Agradecimientos

A mis padres, por darme todo su apoyo incondicional y por cada día confiar y creer en mí y en mis expectativas.

A mis hermanos Raúl y Percy, a mi hermana Nélide y a toda su familia, que de una u otra manera me brindaron su apoyo durante todo el tiempo que duraron mis estudios en Valencia.

A mi tutor Lenin Lemus, por darme la oportunidad de desarrollar este trabajo y un agradecimiento especial a mi tutora Eva Onaindia por su gran dedicación, tiempo, paciencia y las ideas que me fue haciendo llegar para la culminación del presente trabajo.

A *R.* por apoyarme y animarme a continuar siempre hacia adelante.

Resumen

El sistema de acceso a la universidad pública española vincula a dos principales actores: por un lado, las universidades, que son las encargadas de ofertar un cierto número de plazas por cada grado universitario; y, por otro lado, a los estudiantes, quienes tienen la difícil tarea de elegir un grado universitario. El presente estudio tiene como objetivo determinar los factores que influyen en los estudiantes al momento de elegir un determinado grado universitario.

Conocer estos factores es muy importante debido a que ayudaría a las universidades en la oferta de carreras profesionales, y a emprender diversas acciones en pro de la captación del talento adecuado, con el fin de prestigiar a la titulación.

Este estudio se basa en un cuestionario de 25 ítems, respondido por casi 4,400 estudiantes para evaluar diferentes aspectos relacionados con la selección y acceso al sistema universitario español. Con ese cometido, el primer paso fue descubrir las combinaciones lineales que reflejan la mayor variación en los datos del cuestionario a través del análisis de componentes principales (ACP). Luego, se aplicó un análisis factorial exploratorio (AFE) para descubrir la estructura subyacente de los datos originales, cuyo modelo hipotético se aprobó después con un análisis factorial confirmatorio (AFC).

Por último, los resultados mostraron que las 25 preguntas originales del cuestionario podían reducirse a unos cuantos elementos significativos.

Palabras clave: Análisis de Componentes Principales, Análisis Factorial Exploratorio, Análisis Factorial Confirmatorio, Elección de Grado Universitario, Sistema Educativo Español.

Abstract

The process of enrollment in Spain's public universities is linked to two main factors: on the one hand, the universities, which are in charge of offering a certain number of vacancies for each university degree; and, on the other hand, the students, who have the difficult task of choosing a university degree. The study realized aims to determine the factors that influence students' decision when choosing a particular university degree.

Knowing these factors is very important because it would help the universities in choosing which careers to offer, and to undertake various actions to attract the right talent, in order to give prestige to the degree and university.

This study is based on a questionnaire of 25 items, answered by almost 4,400 students to evaluate different aspects related to the selection and access to the Spanish university system. The first step was to discover the linear combinations that reflect the greatest variation in the questionnaire data through principal component analysis (PCA). Then an exploratory factorial analysis (AFE) was applied to discover the underlying structure of the original data, whose hypothetical model was then approved with confirmatory factor analysis (CFA).

Finally, the results showed that the original 25 questions of the questionnaire could be reduced to a few significant elements.

Keywords: Principal Component Analysis, Exploratory Factorial Analysis, Confirmatory Factor Analysis, Spain's Education System, Choosing a University Degree.

Índice de contenidos

1. Introducción	10
1.1 Objetivos	11
1.2 Estructura del documento	12
2. Antecedentes.....	13
2.1 Modelo estructural.....	13
2.2 Cuestionario.....	16
3. Técnicas de análisis de datos	20
3.1 Requerimientos previos.....	20
3.2 Análisis de Componentes Principales.....	21
3.2.1 Matriz de correlaciones y covarianzas.....	22
3.2.2 Criterios para determinar el número de componentes principales	22
3.3 Análisis Factorial	23
3.3.1 Tipos de análisis factorial	24
3.3.2 Métodos de extracción de factores	25
3.3.3 Criterios para determinar el número de factores	25
3.3.4 Rotación e interpretación de los factores	26
3.3.5 Validación del modelo	27
4. Análisis y preparación de los datos	28
4.1 Fuente de datos y características principales	28
4.2 Análisis preliminar de los datos.....	30
5. Aplicación de las técnicas de análisis de datos	33
5.1 Aplicación del análisis de componentes principales	33
5.1.1 Número de componentes principales.....	33
5.1.2 Interpretación de los componentes.....	36
5.2 Análisis factorial	42
5.2.1 Análisis factorial exploratorio (AFE).....	43

5.2.1.1	Número de factores.....	43
5.2.1.2	Interpretación de factores	45
5.2.2	Análisis factorial confirmatorio (AFC)	47
6.	Conclusiones.....	53
7.	Referencias	55
8.	Anexo I	59



Índice de figuras

Figura 1. Diagrama del modelo estructural.....	15
Figura 2. Diagrama de caminos del modelo de medidas del cuestionario para la valoración de los factores	18
Figura 3. Resultado del Test de Bartlett.....	31
Figura 4. Resultado del cálculo de índice de Kaiser-Meyer-Olkin.....	32

Índice de tablas

Tabla 1. Resultados de ACP con rotación oblicua	36
Tabla 2. Resultados de AFE con rotación oblicua	44
Tabla 3. Definición del primer modelo.....	48
Tabla 4. Estadísticos de ajuste del (M1)	48
Tabla 5. Definición del segundo modelo	50
Tabla 6. Estadísticos de ajuste del (M2).....	50
Tabla 7. Definición del tercer modelo	51
Tabla 8. Estadísticos de ajuste del (M3).....	51
Tabla 9. Estadísticos de ajuste para los modelos de AFC.....	51
Tabla 10. Matriz ACP con 25 componentes (parte I)	59
Tabla 11. Matriz ACP con 25 componentes (parte II).....	60
Tabla 12. Matriz ACP con 25 componentes (parte III).....	60
Tabla 13. Segunda matriz de ACP	61



CAPÍTULO I

Introducción

En el mundo entero, la elección de un grado universitario es una de las decisiones más trascendentales para un estudiante, pues supone el comienzo de su carrera profesional y por consiguiente de su proyecto de vida. Es por ello que este paso tan importante, en ocasiones, puede originar ansiedad e indecisión por la gran importancia de la elección, la corta experiencia del alumno y, en muchos casos, por la falta de orientación vocacional.

Como en otras partes del mundo, en Europa la gran gama de grados universitarios y lugares donde poder realizar dichos estudios, hacen que sea mucho más difícil la elección de una carrera universitaria; aun cuando al término del bachillerato se espera que los jóvenes tengan un panorama claro respecto a su futuro estudiantil.

En un estudio realizado por la Fundación BBVA, que realizó una encuesta sobre el comportamiento y actitudes de los jóvenes universitarios en seis países europeos respecto a sus estudios y sus perspectivas vitales y profesionales, se evidenció que a la pregunta «cuál es la motivación para matricularse en determinados estudios», la respuesta más frecuente fue que los estudiantes eligen los estudios que son de su preferencia (Fundación BBVA, 2010).

En el caso del acceso al sistema universitario español, las universidades hasta ahora cumplen un papel pasivo porque es la administración estatal competente la encargada de seleccionar a los futuros alumnos. La admisión a cada carrera profesional empieza con la selección de las mayores notas recabadas en cada una de las pruebas de acceso de los estudiantes, culminando el proceso cuando se cubre el número de vacantes ofertado para dicha titulación. Esto podría, sin embargo, cambiar con la entrada en vigor de una nueva regulación que permitiría a las universidades establecer su propio proceso de admisión (Ruiz R., 2012 junio 30).

Por otro lado, la política universitaria española se ha enfocado fundamentalmente en atraer a un mayor número de estudiantes, dejando de lado la realización de un análisis previo sobre los nuevos perfiles y necesidades del alumnado, como preferencias profesionales o capacidades personales, entre otros (Cabrera, Bethencourt, Pérez & Afonso, 2006). A pesar de la existencia de investigaciones que evidencian la alta motivación de la mayoría de estudiantes por las materias que cursan (Elias-Andreu, 2008), hay otros estudios que ponen de manifiesto la incoherencia de las carreras elegidas por los estudiantes, quienes anteponen como razones sus bajos conocimientos sobre los grados universitarios o sus propias preferencias profesionales (Jiménez & Royo, 1993; Fernández, Peña, Viñuela & Torío, 2007).

Pero también existe un elevado número de universitarios para quienes el motivo de escoger un determinado grado es la perspectiva de encontrar buenas salidas profesionales (Lagares, Lagares & Ordaz, 2012); o aquellos cuya elección se basa en los

consejos paternos en vez de un minucioso análisis de las cualidades personales, de las exigencias académicas y de los perfiles profesionales (Vega & García, 2009).

De estos estudios se puede constatar que el análisis de cada caso es muy complejo ya que la elección de una carrera universitaria puede estar condicionada y afectada por diversas circunstancias. En definitiva, la elección de un grado universitario terminará desencadenando ciertas consecuencias, en ocasiones positivas y prometedoras, si la elección resulta del agrado del estudiante; en caso contrario, una mala elección puede desembocar muchas veces en pérdidas en términos económicos y de esfuerzo, para el estudiante y la universidad; además de ocupar plazas en titulaciones que otros estudiantes con una clara preferencia vocacional podrían haber escogido.

Otras de las consecuencias negativas es el abandono de los estudios, un fenómeno reciente y alarmante en el conjunto de las universidades españolas. De hecho, la tasa de abandono figura como indicador de calidad en numerosos modelos de evaluación de la institución universitaria (Cabrera, Bethencourt, Pérez & Afonso, 2006).

Para abordar la problemática de las motivaciones reales de los estudiantes en la elección de un grado universitario, un reciente estudio publicado en (Soriano Jiménez, 2016) propone un modelo multivariante de relaciones estructurales sobre un conjunto de variables relacionadas con factores sociales, factores individuales y datos de contexto que influyen en la elección de un grado universitario. En este estudio, se presenta además una encuesta de 25 preguntas que respondieron unos 5000 estudiantes de distintas universidades y grados y cuyos datos se utilizaron para obtener y validar el modelo estructural (Peró, Soriano, Capilla, Olmos & Hervás, 2015).

El objetivo del presente trabajo es conocer los principales factores que representan la motivación de los estudiantes a la hora de elegir una carrera universitaria, sin asumir ninguna relación estructural previa entre los posibles factores, teniendo como base el análisis de los datos recopilados en la encuesta presentada en (Soriano Jiménez, 2016). A continuación, se exponen los objetivos principales que se pretende lograr con el desarrollo del presente trabajo.

1.1 Objetivos

El objetivo general de este trabajo es identificar los factores principales que influyen en la elección de un grado universitario en los estudiantes del sistema educativo español. Para ello, se aplicarán dos técnicas de análisis de datos, el análisis de componentes principales y análisis factorial, sobre los datos recopilados de las encuestas realizadas a 5000 alumnos y que se recogen en el trabajo de tesis doctoral mencionado anteriormente (Soriano Jiménez, 2016; Peró et al., 2015). Estas dos técnicas se utilizarán para reducir la dimensionalidad de los datos y plantear un modelo de factores, respectivamente.

Los objetivos específicos a alcanzar con los dos análisis mencionados son:

- Analizar el grado de correlación existente entre todas las variables del conjunto de datos de la encuesta para confirmar el uso adecuado de métodos de extracción de factores.
- Determinar el número de factores principales e interpretar las variables que se comportan como indicadores capaces de explicar la elección de un grado universitario.
- Generar un nuevo modelo reducido de encuesta mediante la aplicación de un análisis factorial exploratorio, y validar la estructura factorial resultante mediante un análisis factorial confirmatorio.

1.2 Estructura del documento

Este documento se estructura de la siguiente forma:

- En el Capítulo II se presentan los principales antecedentes relacionados con la identificación de factores asociados a la elección de un grado universitario en el sistema público de educación superior en España.
- En el Capítulo III se abordan las técnicas de reducción de la dimensionalidad de los datos, como el análisis de componentes principales (ACP), el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC), que se utilizarán para el desarrollo de este trabajo.
- En el Capítulo IV se describe la fuente de datos y sus características principales, y se filtran y preparan los datos para realizar pruebas que corroboren su carácter factorizable, y así proceder con el estudio.
- En el Capítulo V se describe el proceso de extracción de los principales factores y se analizan los resultados obtenidos de la aplicación de las diferentes técnicas de reducción de datos descritos en el Capítulo III.
- El Capítulo VI corresponde a las conclusiones de esta investigación, que podrían servir de base para posibles trabajos futuros.

CAPÍTULO II

Antecedentes

En este capítulo se presentan las principales ideas y contribuciones de la tesis doctoral “*Un modelo estructural para el análisis de los factores asociados a la elección de estudios universitarios*”, de Soriano Jiménez (2016), sobre la cual se sustenta este TFM. En dicha investigación se propone un modelo estructural para analizar el acceso al sistema universitario público español (Hervás, Guàrdia, Però, Capilla & Soriano, 2013), y entre los resultados de esta investigación se incluye una recopilación de datos de estudiantes de primer año de grado a través de una encuesta que son la base del presente trabajo (Però et al., 2015).

A continuación, se definirá la estructura general del modelo estructural y el cuestionario que se utilizó para validar el modelo propuesto en la tesis doctoral.

2.1 Modelo estructural

Previo al análisis y explicación del modelo estructural propuesto por Soriano Jiménez (2016), se repasarán algunos conceptos básicos relacionados con el modelo estructural.

Un modelo de ecuaciones estructurales o modelo estructural es un modelo que se construye mediante una técnica de análisis estadístico multivariante denominada ecuaciones estructurales. Esta técnica permite contrastar modelos a través de una representación gráfica de las relaciones causales y no causales existentes entre las variables propuestas (Ruiz, Pardo & San Martín, 2010). Un modelo de ecuaciones estructurales se compone de:

- **Variables observables:** son aquellas que reflejan datos que se perciben o son observables y, por tanto, son medibles. Por ejemplo, el peso o la edad de una persona, etc.
- **Variables latentes:** “es una *entidad* no observada, que se presume subyace a las variables observadas. El ejemplo mejor conocido de una variable latente importante es *inteligencia*” (Kerlinger & Lee, 2001, p.49).

El objetivo de un modelo estructural es establecer mediante un diagrama las relaciones de causa y efecto existentes entre las variables observables y las variables latentes, de un estudio en particular, con el fin de realizar una validación posterior del modelo generado.

De modo particular, el modelo estructural propuesto en Soriano Jiménez (2016), sobre el que se fundamenta el presente trabajo, identifica las siguientes variables:

- **Variables latentes** asociadas con la elección de un grado universitario. El modelo plantea la existencia de dos tipos de factores –sociales e individuales–, que influyen en la elección de un grado universitario.

- **Factores individuales** que se enfocan en los aspectos relacionados con el estudiante y sus preferencias personales. Algunos factores de carácter académico, como haber obtenido una nota alta en ciertas materias de bachillerato, hacen que el estudiante opte por grados universitarios relacionados con esas materias; otros factores se relacionan con la influencia de su entorno próximo familiar, debido a que los padres o hermanos ya habían estudiado esa carrera en una determinada universidad. Otro factor influyente puede ser la existencia de amigos que estén o hayan cursado cierto grado universitario y ello pueda influir como aliciente para su elección.

El entorno escolar no es ajeno a estos factores, ya que en muchas ocasiones compañeros, orientadores o profesores son los que inciden en mayor o menor medida en la elección. Por último, no se puede dejar de lado la localización geográfica, ya que muchas veces la preferencia por estudiar en un lugar alejado de la ciudad de origen o, en un lugar más cercano al hogar, también se cuenta como un factor individual que forma parte de ese proceso de selección.

- **Factores sociales** que se vinculan con la percepción social relacionada con la calidad y el prestigio de la universidad o la relevancia y nivel de internacionalización de los estudios. Dentro de los factores sociales se incluye también las posibilidades laborales que ofrece la titulación dado que en muchas ocasiones se tiene la percepción de que un grado determinado o ser titulado por alguna determinada universidad garantiza mayores posibilidades de acceder a un puesto de trabajo o mejores salarios, o porque sencillamente se cree que existe demanda por la profesión elegida en el sector.

- **Variables observables** que intervienen de forma determinante en el modelo estructural, y que son datos externos o exógenos a las percepciones sociales e individuales. En este grupo se encuentran los datos relativos al valor de la nota de corte de una titulación y a la oferta y demanda de plazas de una determinada titulación en años anteriores al de la postulación.

Las variables latentes y observables que conforman el modelo estructural propuesto por (Soriano Jiménez, 2016; Hervás, et al., 2013) se muestran en la Figura 1.

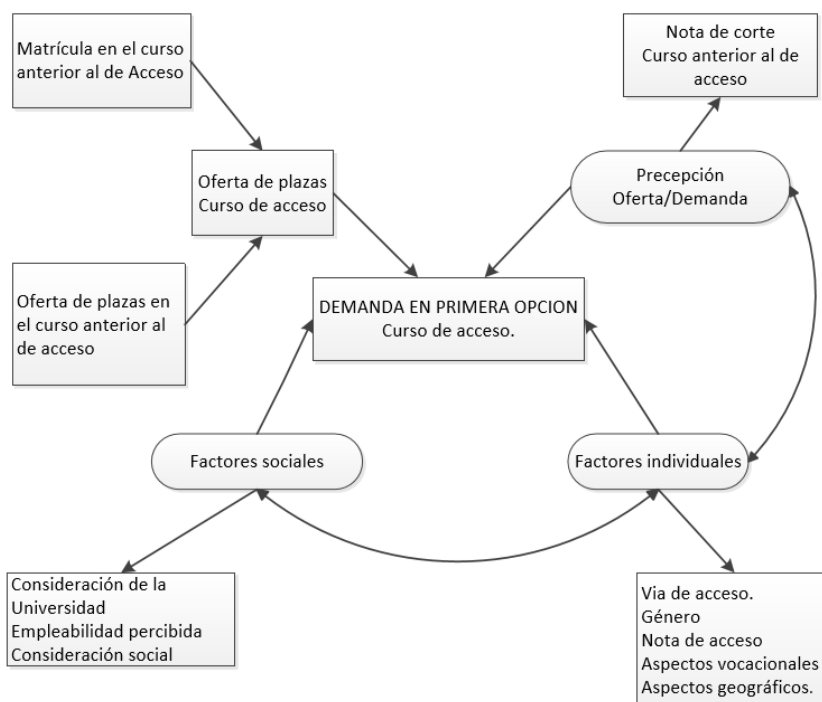


Figura 1. Diagrama del modelo estructural

El diagrama del modelo estructural que se muestra en la Figura 1 se asemeja a un diagrama de flujo que muestra de manera global la relación entre las variables latentes y observables. Los rectángulos representan las variables observables; y las cajas en forma de elipse representan las variables latentes. Las interconexiones existentes entre las variables latentes y las observables se representan a través de flechas que pueden ir en una o ambas direcciones.

Analizando el modelo estructural de la Figura 1 se observa que la variable ‘Oferta de plazas del Curso de acceso’ se encuentra estrechamente relacionada con dos variables observables, ‘Matrícula en el curso anterior al de Acceso’ y ‘Oferta de plazas en el curso anterior al de acceso’; es decir, que la ‘Oferta de plazas del Curso de acceso’ depende directamente del comportamiento de la matrícula en relación con las plazas ofertadas en los años anteriores. La oferta puede incrementarse o reducirse todo dependiendo de la demanda que se haya generado en una titulación en años anteriores.

Por otro lado, la variable latente ‘Percepción Oferta/Demanda’ se define sobre la base de aquello que las personas perciben de la sociedad, de acuerdo con la ‘Nota de corte del Curso anterior al de acceso’. En ocasiones, y debido a que la nota de corte es un indicador conocido, se presta a percepciones subjetivas respecto a la correlación que existe con la demanda de ciertos grados universitarios.

También se puede apreciar que otro condicionante que influye en el proceso de elección de un grado universitario del estudiante son los factores definidos anteriormente como ‘Factores sociales’, los cuales se basan en:

- La ‘Consideración de la Universidad’, donde interviene la percepción de la calidad y prestigio de la institución.
- La ‘Empleabilidad percibida’, que muchas veces no es acorde con la realidad, ya que son percepciones subjetivas pero determinantes para la selección de la universidad o grado. En este sentido, existe la percepción generalizada de que determinados estudios permiten acceder a mejores salarios o tienen una mayor demanda en el mercado laboral.
- Por último, la ‘Consideración social de la titulación’, se relaciona por lo general con el mayor reconocimiento o prestigio a nivel social o internacional de una determinada titulación.

El segundo grupo de factores condicionantes observado en el diagrama del modelo estructural de la Figura 1, son los llamados ‘Factores individuales’, que están conformados por:

- ‘Aspectos vocacionales’, relacionados sobre todo con los gustos o preferencias personales de cada estudiante.
- ‘Aspectos geográficos’, relativos a la elección de los grados en universidades más cercanas al hogar o, todo lo contrario, lo más lejano posible.
- ‘Entorno próximo’, muy relacionado con la influencia que ejerce el entorno próximo familiar (padres, hermanos y amigos) y escolar (tutores, profesores, etc.), en la elección de un grado universitario.
- La vía de acceso en el bachillerato.
- Sexo.
- La Nota de acceso a la universidad.

2.2 Cuestionario

Teniendo en cuenta lo anterior, el modelo estructural propuesto por (Soriano Jiménez, 2016; Perú et al., 2015) se utilizó como guía para la construcción de un cuestionario que se emplearía para la recopilación de datos. El cuestionario se compone de preguntas que reflejan distintos factores sociales e individuales que pueden condicionar la elección de un grado universitario por parte de los estudiantes (ver Sección 2.1).

El primer cuestionario que se diseñó estaba compuesto por 59 preguntas en la escala Likert de 7 puntos, donde las respuestas varían desde 1 (Nada de acuerdo) hasta 7 (Muy de acuerdo). Las preguntas del cuestionario se clasificaban en seis categorías:

- preguntas relacionadas con la Consideración de la universidad (12 ítems).
- preguntas relacionadas con la empleabilidad percibida (10 ítems).
- preguntas relacionadas con la consideración social (9 ítems).
- preguntas relacionadas con los aspectos vocacionales (14 ítems).
- preguntas relacionadas con la influencia del entorno próximo (9 ítems).
- preguntas relacionadas con la localización geográfica (5 ítems).

El cuestionario inicial fue revisado por ocho expertos en el sistema de acceso a la universidad pública española, básicamente para establecer la relación entre cada pregunta o ítem con uno de las seis categorías propuestas, y evaluar la escala de dificultad de respuesta, así como la comprensión y redacción de cada pregunta. Después de ello, se decidió retirar aquellos ítems que no habían sido consignados en ninguno de las categorías propuestas, así como las preguntas consideradas de difícil comprensión. De esa manera, se obtuvo una segunda versión del cuestionario, conformado por 49 ítems en la misma escala Likert, conservando las seis categorías, pero con la siguiente distribución de ítems: Consideración de la universidad (10 ítems), empleabilidad percibida (9 ítems), consideración social (8 ítems), aspectos vocacionales (10 ítems), influencia del entorno próximo (8 ítems) y localización geográfica (4 ítems).

Posteriormente, con la nueva distribución de ítems, se aplicó el cuestionario a una muestra piloto constituida por 252 estudiantes de tres universidades españolas con el objetivo de obtener referencias sobre la validez y fiabilidad del cuestionario. También, mediante un análisis del coeficiente de discriminabilidad de cada uno de los ítems, a efectos de valorar el impacto de su exclusión del cuestionario definitivo, se eliminaron aquellos ítems que poseían un valor de coeficiente de discriminabilidad menor a 0.40, ganándose con ello el incremento de la fiabilidad de la categoría. Asimismo, los datos obtenidos en esta etapa se emplearon para realizar un análisis psicométrico del cuestionario mediante la aplicación de un análisis factorial confirmatorio, obteniéndose valores de fiabilidad y de validez para cada una de las categorías propuestas, que permitieron establecer la bondad del cuestionario.

Para finalizar, tras la realización del análisis psicométrico y de discriminabilidad de cada uno de los ítems en la segunda versión del cuestionario, se generó la versión final del mismo, compuesto por 25 preguntas distribuidas de la siguiente manera: Consideración de la universidad (5 ítems), empleabilidad percibida (3 ítems), consideración social (3 ítems), aspectos vocacionales (4 ítems), influencia del entorno próximo (8 ítems) y localización geográfica (2 ítems).

En la Figura 2, se puede apreciar la distribución de cada uno de los ítems que conforman el cuestionario definitivo. Estos se encuentran organizados de tal manera que permiten evaluar los factores relacionados con los aspectos sociales y personales que intervienen en el proceso de elección de un grado universitario en el sistema público español.

Cabe resaltar que el referido diagrama representa la distribución del cuestionario final, del cual se pretende reducir la cantidad de ítems que lo conforman, para obtener un nuevo cuestionario con menos ítems y, así, tener un instrumento de medición más sencillo y rápido de responder.

Las 25 preguntas que conforman el cuestionario están organizadas de manera general, de acuerdo con los siguientes factores:

a) Factores sociales:

- Elegí entre las carreras profesionales en que sabía que podía ser admitido.
- Prefería estudiar en otro lugar (ciudad, comunidad...).
- La calidad y el prestigio de la universidad fueron determinantes en mi elección.
- Le di más importancia a la carrera profesional que a la universidad.

- El hecho de ser titulado por una universidad u otra da mayores posibilidades laborales.
- La profesión está prestigiada socialmente.
- La profesión es reconocida a nivel internacional.
- Da acceso a una profesión reconocida.
- Creo que los salarios que se consiguen en esta titulación son mejores que en otras.

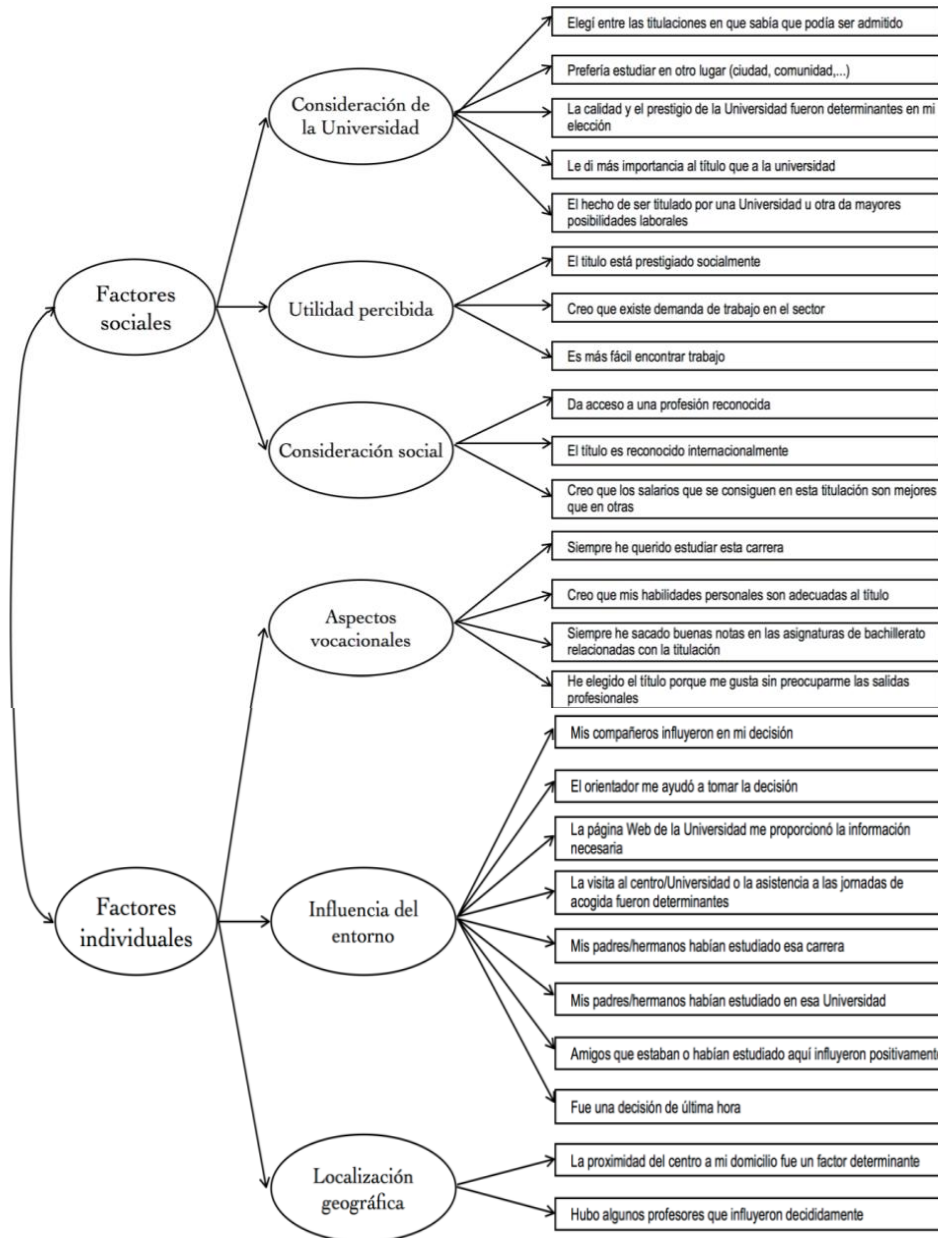


Figura 2. Diagrama de caminos del modelo de medidas del cuestionario para la valoración de los factores

- Creo que existe demanda de trabajo en el sector.
- Es más fácil encontrar trabajo.

b) Factores individuales:

- La proximidad del centro a mi domicilio fue un factor determinante.
- Mis compañeros influyeron en mi decisión.
- El orientador me ayudó a tomar la decisión.
- Hubo algunos profesores que influyeron de modo decidido.
- La página web de la universidad me proporcionó la información necesaria.
- La visita al centro/universidad o la asistencia a las jornadas de acogida fueron determinantes.
- Mis padres/hermanos habían estudiado esa carrera.
- Mis padres/hermanos habían estudiado en esa universidad.
- Amigos que estaban o habían estudiado aquí influyeron en forma positiva.
- Siempre he querido estudiar esta carrera.
- Fue una decisión de última hora.
- Creo que mis habilidades personales se adecúan a la carrera profesional.
- Siempre he sacado buenas notas en las asignaturas de bachillerato relacionadas con la carrera profesional.
- He elegido la carrera profesional porque me gusta, sin preocuparme por las salidas profesionales.

Finalmente, para tener una visión global del estudio se aplicó el cuestionario final a nivel nacional a todo el sistema universitario público español. Para ello, se realizó una encuesta a los alumnos de primer ciclo correspondientes a 116 carreras profesionales de 31 universidades públicas españolas diferentes, obteniéndose una muestra final de 5.111 tras la depuración de datos. El tipo de muestreo utilizado fue accidental y el 76% de información se obtuvo a través de un formulario web *online*; y un 24%, en papel. Los resultados de este cuestionario se utilizaron para validar el modelo estructural.

El conjunto de datos obtenidos como resultado de la encuesta realizada es el que se empleará para el desarrollo del presente trabajo, que tiene como objetivo conocer los principales factores que representan la motivación de los estudiantes a la hora de elegir una carrera universitaria, sin asumir ninguna relación estructural previa entre las posibles categorías.

CAPÍTULO III

Técnicas de análisis de datos

En este capítulo se describen las técnicas de análisis de datos utilizadas en este trabajo. Se emplearon principalmente dos técnicas de análisis estadístico multivariante: el Análisis de Componentes Principales (ACP) y el Análisis Factorial (AF).

Los resultados y la interpretación de los resultados del análisis de los datos con estas dos técnicas se detallan en el Capítulo V.

3.1 Requerimientos previos

Para el procesamiento de los datos, tanto con ACP como AF, existen ciertos cálculos previos a realizar, los cuales indicarán si es pertinente, desde un punto de vista estadístico, llevar a cabo el análisis respectivo, así como la utilización óptima de cada técnica. Una de las principales características que debe cumplir el conjunto de datos a analizar es que exista una alta correlación entre las variables ya que si los datos no están correlacionados no tendría sentido calcular los componentes principales o factores subyacentes en un conjunto amplio de datos.

Bajo este marco, como paso previo, se hace necesario comprobar y examinar la matriz de correlaciones o covarianzas entre todas las variables o ítems objetos de estudio, con el fin de comprobar si los datos en cuestión poseen las características adecuadas para llevar a cabo el análisis. Además de este primer paso exploratorio, existen otras pruebas que permiten evaluar el grado de adecuación de los datos, tal como se detalla en el estudio de Bizquerra (1989; citado por Frías-Navarro & Soler, 2013), donde se aplica en primer lugar la prueba de esfericidad de Bartlett; y posteriormente el índice de Kaiser-Meyer-Olkin (KMO). A continuación, se detalla el contenido de cada una de estas pruebas:

- La **prueba de esfericidad de Bartlett** evalúa la aplicabilidad del análisis de factores; es decir, comprueba si las variables del conjunto de datos a analizar son factorizables (Universidad de Alicante, 2011). Esto se realiza a través de la contrastación de la hipótesis nula de ausencia de correlación entre las variables, lo cual significa que, si se logra rechazar dicha hipótesis, entonces puede afirmarse la existencia de correlación entre las variables (Tabachnick & Fidell, 2001; citado por Frías-Navarro & Soler, 2013). La prueba de esfericidad de Bartlett establece que el modelo estadístico es significativo cuando el valor de fiabilidad resultante (p-valor) es menor de 0.05, y se rechaza cuando el valor de fiabilidad obtenido es mayor a 0.05. Esto quiere decir que se rechaza la hipótesis nula si el p-valor asociado al resultado observado es igual o menor que el nivel de significación establecido, convencionalmente 0.05. Es decir, el p-valor nos muestra la probabilidad de obtener

el resultado observado si suponemos que la hipótesis nula es cierta. En el caso de que el valor sea mayor de 0.05, esto indicaría que no es adecuado realizar una factorización de las variables que forman parte del conjunto de datos en estudio (Universidad de Alicante, 2011).

- **El índice de Kaiser-Meyer-Olkin (KMO)** cumple la misma función que el test de Bartlett, esto es, analizar el grado de relación conjunta de las variables y la adecuación de la muestra. Los valores del índice KMO fluctúan entre 0 y 1; este índice indica que el análisis a realizar será más adecuado cuando el valor que alcance sea lo más próximo a 1, lo cual significa que existe una mayor relación entre los ítems o variables del estudio. Por el contrario, si el índice es bajo, próximo a 0, no sería adecuado aplicar una factorización, ya que el resultado que se obtendría carecería de relevancia (Universidad de Alicante, 2011; Carmona, 2014).

3.2 Análisis de Componentes Principales

El análisis de componentes principales (ACP) fue creado por Karl Pearson en 1901, pero desarrollado y aplicado por Harold Hotelling en 1933. La importancia y utilidad de esta técnica creció gracias al desarrollo de las aplicaciones informáticas. En la actualidad, es quizás una de las técnicas estadísticas multivariantes más populares y utilizada por casi todas las disciplinas científicas (Abdi & Williams, 2010).

El propósito que persigue un ACP es reducir el tamaño de un gran conjunto de datos a un número mucho menor, conservando la mayor información posible del conjunto de datos. Una efectiva reducción de la dimensionalidad de los datos se logra siempre y cuando estos posean variables con una alta correlación.

La reducción de la dimensionalidad de los datos se realiza a través de combinaciones lineales de las variables originales analizadas. El resultado del ACP es un nuevo conjunto de datos no correlacionados, sin información redundante, que se denominan componentes principales, ordenados de tal forma que las primeras componentes conservan la mayor parte de varianza de todas las variables originales; es decir, las primeras componentes representan la mayor cantidad de información respecto a todo el conjunto de datos (Jolliffe, 2002).

En conclusión, con la aplicación del ACP se logra una síntesis de la información inicial, constituida por un gran volumen de datos, mediante la creación de nuevos indicadores, representados mediante componentes principales.

Estos componentes en muchos casos pueden ser fácilmente interpretados y etiquetados por el investigador tras observar los pesos que cada una de las variables aporta a cada uno de los componentes obtenidos. Por ejemplo, si se analizan variables relacionadas con la contaminación atmosférica y se obtiene que las variables referentes a la temperatura y la cantidad de lluvia se agrupan en una componente principal, dicha componente podría asociarse con datos relativos al tipo de clima (De la Fuente Fernández, 2011b).



Por último, cabe resaltar que es posible utilizar el ACP como un paso previo al empleo de otras técnicas, tal como se ha aplicado en este trabajo.

3.2.1 Matriz de correlaciones y covarianzas

Mediante estas dos matrices es posible realizar el cálculo de componentes principales (González, Díaz de Pascual, Torres & Garnica, 1994; Cuadras, 2008; Martín, De Pascual, Lezama & Olmos, 1994):

- **Matriz de correlaciones.** Habitualmente, se calculan los componentes sobre variables originales estandarizadas, es decir, variables con media 0 y varianza 1. Esto equivale a calcular los componentes principales de los datos a través de la matriz de correlaciones (en las variables estandarizadas coinciden las covarianzas y las correlaciones).

La matriz de correlaciones se utiliza generalmente cuando las variables representan conceptos en distintas unidades de medida, ya que tiende a eliminar los efectos distorsivos de datos medidos en distintas unidades.

Adicionalmente, también se utiliza este tipo de matriz cuando se quiere dar la misma importancia a todas y a cada una de las variables del estudio. Esto puede ser conveniente cuando el investigador considera que todas las variables son igualmente relevantes.

- **Matriz de covarianzas.** Se utiliza cuando las variables no están estandarizadas. Generalmente, conviene utilizar este tipo de matriz cuando todas las variables poseen las mismas unidades de medida. Por ejemplo, si se analizan datos referentes al clima y todas las variables a analizar son relativas a la temperatura y se encuentran medidas en grados centígrados.

También se sugiere emplear este método cuando el investigador considera que es más pertinente destacar cada una de las variables en función de su grado de variabilidad.

Por último, cuando todas las variables están representadas en las mismas unidades, se puede utilizar tanto la matriz de correlaciones como la matriz de covarianzas; no obstante, en caso de duda, es conveniente realizar ambos análisis, y seleccionar aquel que genere interpretaciones más informativas y facilite el proceso de análisis.

3.2.2 Criterios para determinar el número de componentes principales

En cuanto al número de factores que se deben extraer o seleccionar, existe una gran variedad de criterios, de los cuales solo se explicarán tres de los más utilizados en la actualidad (Cuadras, 2008).

- **Criterio de porcentaje de varianza.** Este criterio plantea conservar aquellos componentes principales cuyos porcentajes acumulados aporten la mayor variabilidad explicada. A veces, este criterio puede tener ciertos inconvenientes a la hora de realizar la selección de los componentes principales, sobre todo cuando se analiza un conjunto de datos de más de 50 ítems debido a que puede conducir a la extracción de demasiados componentes. Cuando se utiliza este criterio, es recomendable conservar aquellos componentes cuyos porcentajes de varianza sumados en total expliquen al menos un 50% de la variabilidad total (Merenda, 1997; citado por Pérez & Medrano, 2010).
- **Criterio de Kaiser.** Tal y como se ha comentado en el punto 3.2.1, la matriz de correlaciones se aplica cuando las variables están estandarizadas, es decir variables que siguen una distribución normal $N(0,1)$. En este caso, dado que la varianza de las variables es 1, se puede afirmar que la componente principal cuya varianza sea menor a 1 explicará menos variabilidad que las otras variables. El criterio de Kaiser se basa en mantener aquellas componentes cuyas varianzas sean mayores a la unidad (Cuadras, 2008). La razón por que se fija este punto de corte se debe a que como las variables están tipificadas, resultaría inadecuado interpretar un componente que explique una varianza menor de 1 en comparación con la explicada por una variable en particular, según Kahn (2006; citado por Pérez & Medrano, 2010).
- **Scree test o gráfica scree.** Este criterio de extracción de componentes principales es uno de los más utilizados en la actualidad. La gráfica scree es una representación gráfica compuesta por los llamados autovalores o *loadings* en el eje vertical, de acuerdo al orden de su magnitud descendente; y el número de factores o componentes principales en el eje horizontal. Esta distribución permite identificar el número óptimo de factores o componentes a extraer. La ruptura entre la fuerte pendiente que se origina desde el eje vertical y la estabilización de la misma, indicará el número de factores significativos a extraer (Cattell, 1966).

3.3 Análisis Factorial

El análisis factorial (AF) es una técnica estadística que se centra en explicar la relación de variables correlacionadas y observadas con respecto a un conjunto de factores (Kline, 2014). Su objetivo es encontrar la variación conjunta entre las variables latentes no observadas y la respuesta de las variables observadas. En el AF las variables observadas se definen como combinaciones lineales de los factores. La ganancia de información sobre la interdependencia de los datos observados se utiliza para reducir el número de variables observadas del conjunto de datos. Este análisis se utiliza para tratar un conjunto de datos donde un número significativo de observaciones se cree que aportan la misma información en base a un pequeño número de variables subyacentes. En otras palabras, simplifica conjuntos complejos de datos.



Conceptualmente, el AF presenta dos tipos o modalidades diferentes: el análisis factorial exploratorio (AFE) y el análisis factorial confirmatorio (AFC).

3.3.1 Tipos de análisis factorial

Los tipos de AF más utilizados son los siguientes:

- El análisis factorial exploratorio (AFE) mide básicamente los factores que afectan las variables en una estructura de datos dada. La lógica en la cual se basa el AFE radica en que, si en un conjunto de datos existen altas correlaciones entre sí, eso se debe a que tienen factores en común (Pérez y Medrano, 2010; Blaclock, 1966). En otras palabras, el AFE intenta encontrar la estructura latente presente en los datos muestrales (Pérez-Gil, Moscoso & Rodríguez, 2000; Bollen, 1989).

La aplicación del AFE parte del supuesto de que las variables observadas forman parte de cierto número de factores o variables latentes comunes. Esto queda evidenciado tras la aplicación del AFE, el cual devolverá una matriz factorial conformado por un cierto número de factores y las respectivas cargas factoriales de cada una de las variables observadas, las cuales permitirán posteriormente la catalogación de cada uno de los factores (Pérez-Gil et al., 2000).

Por lo general, el AFE se utiliza con el fin de crear estructuras de los modelos teóricos e hipótesis que puedan contrastarse a nivel empírico (Gorsuch, 2010; citado por Pérez-Gil et al., 2000).

- El análisis factorial confirmatorio (AFC) es una técnica estadística que se emplea para evaluar la significación estadística y el ajuste de un modelo teórico de factores que representan las variables observadas originales.

La aplicación del AFC se inicia con un modelo previamente definido. Este modelo se puede construir por medio de la experimentación y/o la observación. En el presente trabajo dicho modelo se genera a partir del resultado de la aplicación de un AFE previo.

Una vez que se tiene la definición del modelo, se procede a estimar los indicadores, los cuales permitirán medir el ajuste del modelo. Si el modelo presenta un ajuste adecuado, se podrá utilizar para evaluar e interpretar los indicadores. De lo contrario, se modificará el modelo y de nuevo se iniciará el proceso de validación (Fadlelmula, 2011; citado por Morata-Ramírez, Holgado-Tello, Barbero-García & Mendez, 2015).

La diferencia más marcada frente al uso tanto del AFE y el AFC radica en que este último se conduce principalmente por teorías sustantivas y por expectativas, mientras que el primero se basa en datos que intentan descubrir la estructura subyacente que estos poseen (Bollen, 1989; citado por Pérez-Gil et al., 2000).

3.3.2 Métodos de extracción de factores

Dentro del análisis factorial se contempla una gran variedad de métodos que permiten la extracción de factores, entre los que se encuentran:

- **Método de máxima verosimilitud:** conocido también como *Maximum Likelihood Estimation* (MLE), su principal objetivo es estimar la significación estadística de los pesos factoriales y generar intervalos de confianza de los mismos. Proporciona estimadores insesgados, consistentes, eficientes y normalmente distribuidos. Además, goza del respaldo de muchos autores quienes sugieren que es uno de los métodos más adecuados cuando se trabaja con datos que poseen una distribución normal (Byrne, 2001; Costello & Osborne, 2005; citado por Pérez & Medrano, 2010).
- **Método de los ejes principales:** es un método de estimación recursivo que estima la varianza que las variables tienen en común. Se calcula mediante la sustitución de una matriz de correlaciones poblacional por una matriz de correlaciones muestral, consiste en combinar de manera iterativa una estimación de la matriz de especificidades (representa la parte de la varianza específica de cada variable) con una estimación de la matriz de cargas factoriales hasta que los valores dichas estimaciones apenas cambien. Se recomienda el uso de este método cuando se trabaja con datos que no cumplen el supuesto de normalidad multivariada (De la Fuente Fernández, 2011).

3.3.3 Criterios para determinar el número de factores

Además de los criterios ya descritos en el punto 3.2.2, los cuales también pueden aplicarse en la determinación del número de factores a extraer, existen algunos criterios exclusivos del análisis factorial, que se describen a continuación:

- **Criterio de división a la mitad:** consiste en realizar un análisis factorial a dos muestras, las cuales deberán ser extraídas de manera aleatoria y donde cada una representa la mitad de la muestra principal. De este análisis solo se seleccionarán aquellos factores que posean los valores más altos de cada uno de los resultados obtenidos (De la Fuente Fernández, 2011).
- **Análisis paralelo:** este método consiste en crear autovalores de una matriz de datos aleatorios con el mismo número de variables y el mismo tamaño una muestra que la matriz de datos original. La selección de los factores se realiza comparando el autovalor de cada factor generado en la matriz de datos aleatorios y la matriz de los datos originales, aquellos autovalores de los datos originales superiores a los autovalores de los datos aleatorios nos indicarán el número de factores óptimos que se deben extraer (Horn, 1965).



3.3.4 Rotación e interpretación de los factores

El resultado final que se obtiene al aplicar un ACP o AFE es una matriz factorial, a cada elemento de esta matriz se le denomina saturaciones, ponderación, pesos o cargas factoriales de cada una de las variables del conjunto de datos analizados con los componentes o factores obtenidos.

Por lo general, la matriz inicial resultante suele generar ciertas dificultades al investigador durante la interpretación de los resultados, sobre todo cuando la solución inicial da como resultado la extracción de más de un componente principal o factor.

La interpretación de los factores o componentes principales se realiza analizando los pesos factoriales que las variables observadas aportan a cada uno de los factores o componentes principales obtenidos, de tal forma que las variables con mayor peso en cada uno de los factores o componentes principales contribuirán a poder asociar un nombre a cada uno de los factores o componentes principales teniendo como ayuda el conocimiento teórico sobre las variables en estudio.

El objetivo de realizar una rotación es poder obtener una solución que facilite su interpretación; lo que se consigue con su aplicación es eliminar las correlaciones negativas y reducir el número de contribuciones de los ítems en cada uno de los componentes principales o factores. En otras palabras, la rotación se emplea para que las cargas factoriales de los ítems sean lo más cercana a 1 con uno de los componentes o factores y cargas factoriales próximas a 0 con el resto de los demás componentes o factores.

Existen dos tipos de rotación que pueden aplicarse en tal caso: el de tipo ortogonal y el oblicuo (De la Fuente Fernández, 2011).

La ventaja principal de las rotaciones ortogonales es su sencillez, ya que los pesos factoriales representan las contribuciones de las variables en cada uno de los componentes o factores. Entre las rotaciones ortogonales que más se utilizan son la *Varimax* (disminuye el número de variables con pesos factoriales altos en un factor o componente principal), la *Quartimax* (reduce el número de factores o componentes principales necesarios para interpretar una variable) y la *Equamax* que es la combinación de las dos rotaciones anteriores.

En el caso de la rotación oblicua, que es considerada más realista que la ortogonal, suele generar resultados más interpretables y también aporta información sobre la relación entre los factores. Los pesos factoriales no coinciden con las correlaciones entre los factores o componente y la variable, debido a que los factores o componentes están correlacionados entre sí. El método más utilizado es el *Oblimín*, el cual considera como parámetro de control el grado de correlación entre factores o componentes y las variables, con valores preferentemente entre -0.5 y 0.5 .

3.3.5 Validación del modelo

El último paso al realizar un AFC es estudiar la validez del modelo generado a través del AFE. Por tal motivo se han establecido los siguientes cinco índices o indicadores de ajuste que permiten medir la validez del modelo.

- **El Chi-cuadrado:** es una función de las diferencias entre las covarianzas observadas y las covarianzas implicadas por el modelo. Se ha sugerido que un Chi-cuadrado dos o tres veces mayor que los grados de libertad es aceptable (Carmines & McIver, 1981).
- **El chi-cuadrado relativo:** también llamado chi-cuadrado normal, es el índice de ajuste chi-cuadrado dividido por grados de libertad, en un intento de hacer que el chi-cuadrado sea menos dependiente del tamaño de la muestra (Chi-cuadrado/grados de libertad). Valores entre [1,2] o [1,3] indican un ajuste aceptable entre el modelo hipotético y los datos de muestra (Carmines & McIver, 1981). Diferentes investigadores han recomendado usar una proporción tan baja como 2 o tan alta como 5 para indicar un ajuste razonable (Marsh & Hocevar, 1985).
- **CFI** (índice comparativo de corrección) y **TLI** (índice de Tucker-Lewis), cercano a 1 indica un ajuste muy bueno; si el valor es mayor de 0.9 o cercano a 0.95 indica buen ajuste. Por convención, CFI debe ser igual o superior a 0.90 para aceptar el modelo (Hu & Bentler, 1999).
- **RMSEA** (error cuadrático medio cuadrado de aproximación); algunos estudios sugieren un valor de RMSEA \leq .06 como el punto de corte para denotar un adecuado ajuste (Hu & Bentler, 1999). A medida que este valor decrece el ajuste aumenta y adopta el valor de 0 ante un ajuste perfecto, de 0.08 a 0.10 indica un ajuste mediocre, mayor a 0.10 indica un mal ajuste.
- **SRMR** (índice de la raíz de los cuadrados medios del residuo), este indicador es la raíz cuadrada de las diferencias entre las varianzas reales y las covarianzas y las varianzas y covarianzas generadas asumiendo que el modelo es verdadero. Un valor menor a 0.08 significa un buen ajuste, cuanto menor es el SRMR, mejor es el ajuste del modelo, de manera que un valor igual a 0 indica un ajuste perfecto (Hu y Bentler, 1999).



CAPÍTULO IV

Análisis y preparación de los datos

En esta sección se describe el conjunto de datos que se ha utilizado para realizar el análisis estadístico junto con sus principales características. También se detalla el proceso de preparación de los datos y se presenta un análisis preliminar para confirmar que el uso de las técnicas ACP y AF son apropiadas para este conjunto de datos.

4.1 Fuente de datos y características principales

El conjunto de datos a utilizar en el análisis de este trabajo proviene del resultado obtenido en la tesis doctoral de (Soriano Jiménez, 2016). Tal como se mencionó en la sección 2.2, en dicha tesis se realizó una encuesta a alumnos de primer año de grado de un total de 116 titulaciones de 31 universidades pertenecientes al sistema universitario público español. La muestra final contenía un total de 4,364 encuestas, después de la depuración realizada a todo el conjunto de datos. Cabe resaltar que esta muestra está compuesta únicamente por alumnos provenientes de bachillerato, ya que se decidió excluir a los estudiantes de Ciclos Formativos de Grado superior y otros estudios previos los cuales no representaban una muestra sustancial en todo el conjunto de datos a analizar.

Tal y como se comentó en la sección 2.2, se trata de un cuestionario compuesto de 25 preguntas en la escala Likert de 7 puntos. Este tipo de escala es muy utilizado en la recolección de datos en investigaciones de tipo social para medir actitudes. La escala de Likert está compuesta por un conjunto de ítems bajo la forma de afirmaciones o juicios ante los cuales se solicita la reacción (favorable o desfavorable, positiva o negativa) de los individuos. En el caso que nos ocupa se utilizó la escala Likert de 7 niveles, donde las respuestas van desde 1 (Nada de acuerdo) hasta 7 (Muy de acuerdo).

A continuación se vuelve a detallar las 25 preguntas que conforman el cuestionario final, el cual se explicó detalladamente en la sección 2.2. En este capítulo mostramos las 25 preguntas junto con un identificador asociado de la forma <p_número> que es el que se ha utilizado en la codificación de las pruebas y el que se utilizará para presentar los resultados finales en cada análisis a realizar.

- p_12** Elegí entre las titulaciones en que sabía que podía ser admitido.
- p_13** La proximidad del centro a mi domicilio fue un factor determinante.
- p_14** Prefería estudiar en otro lugar (ciudad, comunidad, ...)

- p_15** Mis compañeros influyeron en mi decisión.
- p_16** El orientador me ayudó a tomar la decisión.
- p_17** Hubo algunos profesores que influyeron decididamente.
- p_18** La página Web de la Universidad me proporcionó la información necesaria.
- p_19** La visita al centro/Universidad o la asistencia a las jornadas de acogida fueron determinantes.
- p_20** Mis padres/hermanos habían estudiado esa carrera.
- p_21** Mis padres/hermanos habían estudiado en esa Universidad.
- p_22** Amigos que estaban o habían estudiado aquí influyeron positivamente.
- p_23** Siempre he querido estudiar esta carrera.
- p_24** Fue una decisión de última hora.
- p_25** Creo que mis habilidades personales son adecuadas al título.
- p_26** Siempre he sacado buenas notas en las asignaturas de bachillerato relacionadas con la titulación.
- p_29** La calidad y el prestigio de la Universidad fueron determinantes en mi elección.
- p_30** Le di más importancia al título que a la universidad.
- p_31** El hecho de ser titulado por una Universidad u otra da mayores posibilidades laborales.
- p_32** El título está prestigiado socialmente.
- p_33** El título es reconocido internacionalmente.
- p_34** Da acceso a una profesión reconocida.
- p_36** He elegido el título porque me gusta sin preocuparme las salidas profesionales.
- p_37** Creo que los salarios que se consiguen en esta titulación son mejores que en otras.
- p_38** Creo que existe demanda de trabajo en el sector.
- p_39** Es más fácil encontrar trabajo.

4.2 Análisis preliminar de los datos

Como se ha podido ver en la sección 3.1, hay una serie de requerimientos previos que debe cumplir el conjunto de datos a analizar para garantizar que la aplicación del ACP y AF devolverá resultados estadísticamente significativos y susceptibles de interpretación sustantiva.

En primer lugar, debe garantizarse que la aplicación de un análisis factorial se realiza sobre variables cuantitativas. En nuestro caso estamos trabajando con una escala Likert que cuenta con opciones de respuesta ordinal en 7 categorías. Dicha distribución tiene propiedades que permiten poder considerarlas como variables cuantitativas, tal y como muestran diversos estudios que determinan que la utilización de escalas a partir de 6 categorías de respuesta estabiliza la consistencia interna de los datos recogidos (Gelin, Beasley & Zumbo, 2003), (Lozano, García-Cueto & Muñiz, 2008), (Ramsay, 1973), (Weng, 2004), (Ferrando & Lorenzo-Seva, 2014).

Por otro lado, en la investigación llevada a cabo en la tesis doctoral (Soriano Jiménez, 2016) se constata que la distribución de los datos tiene propiedades que permiten usar valores agregados como variables continuas.

Antes de realizar el análisis de factores de los datos recogidos del cuestionario, es necesario analizar el grado de correlación de los 25 ítems en estudio, calculando y examinando la matriz de correlaciones de las variables la cual se muestra en el Gráfico 1.

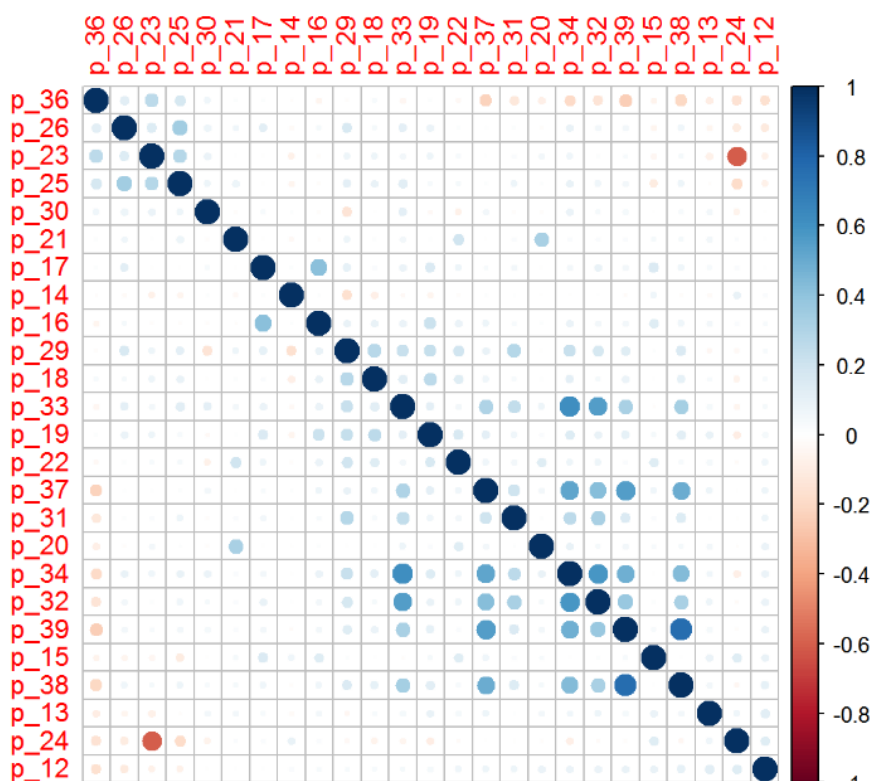


Gráfico 1. Matriz de correlaciones

El Gráfico 1 muestra distintos colores en las celdas dependiendo del nivel de correlación entre dos variables, las cuales se encuentran representadas en las filas y columnas de la matriz. En la parte de la derecha se muestra una barra de gama de colores que representa el grado de correlación entre -1 y 1. El color azul indica una fuerte correlación positiva, y el color rojo indica una correlación negativa. Los colores van bajando su intensidad a medida que la correlación disminuye.

Como podemos ver, el Gráfico 1 permite identificar de manera rápida cuáles son las variables que correlacionan entre sí. A partir de dicho análisis, se puede afirmar que la mayor parte de las variables presentan bajas correlaciones, motivo por el cual es necesario realizar el test de esfericidad de Bartlett y la prueba de Kaiser-Meyer-Olkin (KMO), con objeto de determinar si la aplicación de ACP y AF al conjunto de datos es significativa.

El test de esfericidad de Bartlett es una de las primeras pruebas que se propone realizar como requerimiento previo a la aplicación de técnicas de análisis de datos utilizadas en este trabajo. Una descripción de esta prueba se detalla en la sección 3.1. Para complementar dicha explicación, podemos decir que esta prueba evalúa la hipótesis nula de que no existe correlación entre las variables; es decir, comprueba si la matriz de correlaciones es igual a la matriz identidad. En el caso de que se rechace la hipótesis nula, esto indicaría que sí existe un grado de correlación estadísticamente significativa entre las variables. Al observar los resultados de la prueba de Bartlett en la Figura 3, concretamente el (p.value), observamos que se rechaza la hipótesis nula porque el p-valor que se obtiene es 0, el cual es menor a 0.05. Consecuentemente, podemos afirmar que existe un grado de correlación entre las variables.

```
## $chisq
## [1] 21873.41
##
## $p.value
## [1] 0
##
## $df
## [1] 300
```

Figura 3. Resultado del Test de Bartlett

Por último, la segunda prueba que también fue descrita en la Sección 3.1 es el test de Kaiser-Meyer-Olkin (KMO) o Medida de adecuación de la muestra (MSA), la cual evalúa el grado de correlación entre dos variables a partir de la comparación de las correlaciones parciales y observadas. Esta medida se interpreta a través del valor del índice resultante, el cual toma valores entre 0 y 1. Los valores pequeños indican que el análisis factorial puede no ser una buena idea, dado que las correlaciones entre los pares de variables no pueden ser explicadas por otras variables. Los valores menores a 0.5 indican que no debe utilizarse el análisis factorial con los datos muestrales que se están analizando.

El resultado del índice MSA para los datos de nuestro estudio se muestra en la Figura 4 e indica un valor de 0.75, lo cual significa un nivel de multicolinealidad aceptable entre las variables. Si el valor del indicador MSA hubiera dado un resultado menor a 0.5, habría que considerar un cambio de variables o bien utilizar otra técnica ya que es muy poco probable que funcionen los modelos sin el cumplimiento de esta prueba.

```

## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = newetotal)
## Overall MSA = 0.75
## MSA for each item =
## p_12 p_13 p_14 p_15 p_29 p_16 p_30 p_17 p_31 p_18 p_32 p_19 p_33 p_20 p_34
## 0.74 0.69 0.62 0.65 0.71 0.62 0.59 0.60 0.80 0.72 0.86 0.78 0.81 0.57 0.84
## p_21 p_22 p_36 p_23 p_37 p_24 p_38 p_25 p_39 p_26
## 0.55 0.72 0.77 0.59 0.88 0.59 0.75 0.71 0.75 0.67

```

Figura 4. Resultado del cálculo de índice de Kaiser-Meyer-Olkin

Después de realizar los análisis preliminares sobre los datos de la muestra mediante el análisis de la matriz de correlaciones, el test de esfericidad de Bartlett y la Medida de adecuación de la muestra, todos ellos arrojan resultados favorables. Podemos, por tanto, concluir que nuestro conjunto de datos son factorizables y garantizan que la aplicación de un ACP o un AF devolverá resultados estadísticamente significativos y susceptibles de interpretación sustantiva. En el siguiente capítulo se aplicarán estas técnicas y se describirán los resultados obtenidos con cada una de ellas.

CAPÍTULO V

Aplicación de las técnicas de análisis de datos

En este capítulo se presenta el análisis del conjunto de datos estudio de este trabajo. En primer lugar, se aplicará la técnica denominada ACP con la cual se puede obtener una identificación e interpretación inicial de los componentes principales que explican el conjunto de datos.

En la segunda etapa de este trabajo, se aplica el análisis factorial al conjunto de datos con el objetivo de encontrar un modelo factorial que explique la correlación de las variables respecto a un conjunto de factores. Este análisis se realiza en dos fases. En una primera fase se aplica un análisis factorial exploratorio a la mitad del conjunto de datos el cual permitirá encontrar la estructura latente en dicho conjunto de datos. En una segunda fase se aplica un análisis factorial confirmatorio a la otra mitad del conjunto de datos con el fin de poder verificar la validez del modelo generado.

La aplicación de estas dos técnicas de análisis se realizó con en el paquete estadístico R utilizando las librerías *psych* (W.Revelle & M.Revelle, 2017), (Revelle, 2017) y *lavaan* (Rosseel, 2011).

5.1 Aplicación del análisis de componentes principales

5.1.1 Número de componentes principales

El primer paso es determinar la cantidad de componentes a extraer. Para ello, inicialmente se ejecutó el análisis solicitando tantos componentes como variables involucradas en el análisis, es decir, un total de 25 componentes. En el Anexo 1 se muestran los resultados obtenidos donde se puede observar lo siguiente:

- En la fila 'SS Loadings' de la Tabla 13 se puede observar los autovalores de los componentes extraídos. Estos autovalores proporcionan una medida de tolerancia para poder decidir la cantidad de componentes a retener. Autovalores iguales o mayores a 1 indican que el componente logra explicar más varianza que una variable por sí sola. En este caso, se puede observar que los autovalores de las ocho primeras componentes son mayores o iguales a 1.



- En la fila 'Proportion Var' de la Tabla 13 se puede apreciar la proporción de la varianza que aporta cada componente al total de la varianza acumulada. El primer componente aporta un 16% de variabilidad, el segundo componente aporta un 9% de variabilidad, y así se puede apreciar el porcentaje de variabilidad que va aportando cada componente. Cabe resaltar que a partir de la novena componente la medida de la varianza que explica cada componente es muy baja.
- En la fila 'Cumulative Var' de la Tabla 13 se puede apreciar el porcentaje acumulado de variabilidad que va aportando cada grupo de componentes. Una solución con ocho componentes explicaría un 56% de la varianza total del conjunto de variables en estudio.
- El total de los componentes principales aporta el 100% de variabilidad, pero como el objetivo del ACP es reducir estadísticamente la información original, se debe elegir un número adecuado de componentes principales que tengan a su vez la mayor información posible de las variables originales.

Tras realizar este análisis preliminar, se procedió a aplicar la gráfica scree o *Scree test* para determinar el número de componentes a extraer. Conviene destacar que el *Scree test* es uno de los criterios más empleados para la selección del número de componentes principales, el cual se ha explicado detalladamente en el Capítulo 3.

En el Gráfico 1, podemos apreciar en el eje horizontal el número de componentes y en el eje vertical los autovalores de dichos componentes. Sobre la gráfica se traza una línea recta que atraviese desde la altura de los autovalores de los últimos componentes, los más bajos. Analizando de izquierda a derecha podemos observar que los autovalores que explican la mayor parte de la varianza se encuentran en la parte izquierda formando una fuerte pendiente y el punto de inflexión empieza a ser notorio a partir de la novena componente. Por dicha razón, el número de componentes principales a extraer que sugiere el *Scree test* es 8.

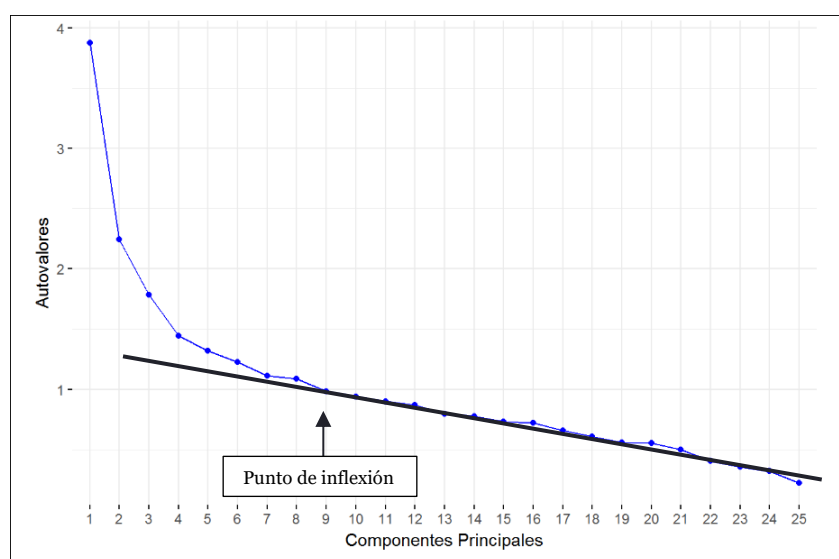


Gráfico 2. Scree Test

Una vez que se conoce el número de componentes a extraer, el siguiente paso es realizar el ACP con ocho componentes, aplicando una rotación oblicua para facilitar la interpretación de cada uno de las componentes resultantes. El resultado de este proceso es la extracción de dos matrices que se muestran en la Tabla 1. Cabe destacar que en dicha tabla no se muestran los pesos factoriales inferiores a 0.3 para una mayor claridad en la interpretación de cada uno de los componentes principales extraídos.

La primera matriz corresponde a la matriz rotada resultante del ACP, Tabla 1 (Matriz 1). En la primera fila de la matriz de componentes se aprecia el nombre de cada componente (PC1, PC2, ..., PC8), mientras que en la primera columna se encuentran las 25 variables codificadas. La intersección entre un elemento de la columna y uno de la fila representa el valor del peso que el ítem correspondiente aporta a cada componente.

Para decidir los ítems que contribuyen a cada componente, se debe observar la Matriz 1 de la Tabla 1, y elegir los ítems con mayores valores de saturación (positivos o negativos). En base a estos valores de saturación se determina la composición de cada componente. Cabe resaltar que como la matriz está rotada, la mayoría de variables contribuyen únicamente a un solo componente, especialmente si el valor de la saturación de la variable en la componente es alto.

En la Matriz 1 de la Tabla 1 se muestra mediante círculos de colores el valor más alto de las variables que contribuyen a cada componente. Así se observa que en la componente 1 (PC1), las variables que tienen las saturaciones más altas son p_38 y p_39, lo que implica que este componente queda definido con estas dos variables. Las variables p_31, p_32 y p_33 tienen las saturaciones más altas en el segundo componente, lo que hace que el segundo componente quede definido con estas tres variables.

Del mismo modo, así sucesivamente podemos observar cómo según las saturaciones más altas de las variables definen a los componentes.

En la segunda matriz de la Tabla 1 también se aprecia los siguientes valores:

- En la fila 'SS Loadings' se muestra el valor de la medida de la varianza que explica cada componente, en este caso los ocho componentes superan la unidad.
- En la fila 'Proportion Var' se puede apreciar la proporción de la varianza que aporta cada componente al total de la varianza acumulada de los ocho componentes.
- En la fila 'Cumulative Var' indica que la extracción de ocho componentes podría representar el 56% de la varianza total de los datos.



```

## Principal Components Analysis
## Call: principal(r = dtotalest, nfactors = 8, rotate = "oblimin", covar = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##      PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  h2  u2  com
## p_12                -0.37      0.31 0.33 0.67 3.3
## p_13                0.66 0.52 0.48 1.6
## p_14                -0.46      0.26 0.74 1.9
## p_15                0.45      0.32 0.68 2.2
## p_29      0.32      0.56      -0.33 0.63 0.37 2.5
## p_16                0.75      0.59 0.41 1.0
## p_30                0.79      0.64 0.54 0.46 1.6
## p_17                0.72      0.64 0.36 1.1
## p_31      0.72      0.50 0.50 1.3
## p_18      0.76      0.69 0.48 0.52 1.1
## p_32      0.76      0.66 0.34 1.1
## p_19      0.72      0.63 0.37 1.2
## p_33      0.72      0.58 0.42 1.1
## p_20      0.72      0.70 0.30 1.6
## p_34  0.34  0.62      0.75      0.58 0.36 1.2
## p_21      0.78      0.64 0.36 1.2
## p_22                0.47      0.41 0.59 2.6
## p_36 -0.37      0.30      0.34      0.40 0.60 3.2
## p_23                0.87      0.78 0.22 1.0
## p_37  0.64      0.59 0.41 1.3
## p_24                -0.87      0.76 0.24 1.0
## p_38  0.88      0.77 0.23 1.0
## p_25                0.64      0.51 0.49 1.3
## p_39  0.89      0.81 0.19 1.0
## p_26                0.76      0.60 0.40 1.1
##
##      PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8
## SS loadings      2.48 2.37 1.77 1.61 1.62 1.58 1.49 1.18
## Proportion Var    0.10 0.09 0.07 0.06 0.06 0.06 0.06 0.05
## Cumulative Var    0.10 0.19 0.26 0.33 0.39 0.46 0.52 0.56
## Proportion Explained 0.18 0.17 0.13 0.11 0.12 0.11 0.11 0.08
## Cumulative Proportion 0.18 0.34 0.47 0.58 0.70 0.81 0.92 1.00

```

Tabla 1. Resultados de ACP con rotación oblicua

Finalmente, después de realizar la extracción de los ocho componentes principales, el siguiente paso es realizar la interpretación de cada uno de ellos.

5.1.2 Interpretación de los componentes

En la Tabla 1 se muestra el resultado de la aplicación del ACP con rotación oblicua al conjunto de datos en estudio. La interpretación de los componentes se realizará en base a los datos de la Matriz 1.

Tal y como se ha comentado, la Matriz 1 representa la saturación de los ítems o preguntas del cuestionario, las cuales aparecen codificadas con un indicador p-x (ver sección 4.1), a cada uno de los ocho componentes, que es el número de componentes a extraer resultante de la aplicación del *Scree test*.

Cada uno de los componentes principales estará entonces definido por los ítems que contribuyen con el mayor peso. De este modo, estos resultados nos permiten asignar posibles denominaciones a cada uno de los componentes principales extraídos.

El **primer componente** combina la facilidad de encontrar trabajo y la creencia de que existe una demanda en el sector, y en menor medida la creencia de que los salarios que se consiguen con cierta titulación son mejores que con otras titulaciones, así como el acceso a una profesión reconocida. Esto se debe a que tanto la variable p_39 como la variable p_38 tienen una alta saturación, cercana a uno, mientras que las variables p_37 y p_34 toman valores más bajos, (ver Gráfico 3). La primera componente principal (PC1) podría definirse como un componente que representa la **Empleabilidad percibida de la titulación o grado** que cursa el estudiante.

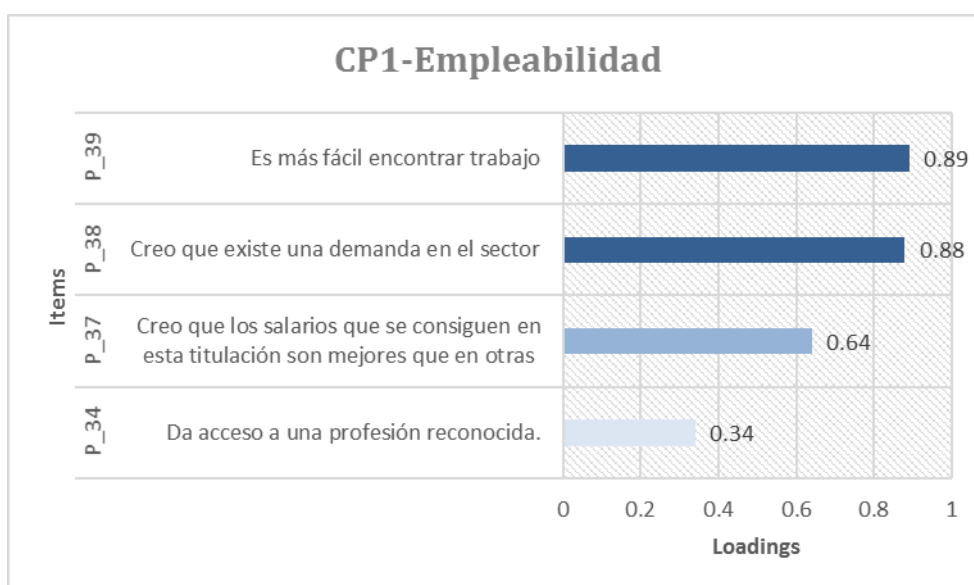


Gráfico 3. Primera componente (PC1)

En la segunda componente principal (PC2), las variables p_32, p_33 y p_31 se encuentran muy correlacionadas entre sí (ver Gráfico 4). Estas tres variables están relacionadas con la percepción social vinculado con el prestigio o la relevancia y nivel de internacionalización del título. Por tal motivo, estas tres variables conforman el **segundo componente** que se define como **Consideración Social de la titulación o grado que cursa el estudiante**.

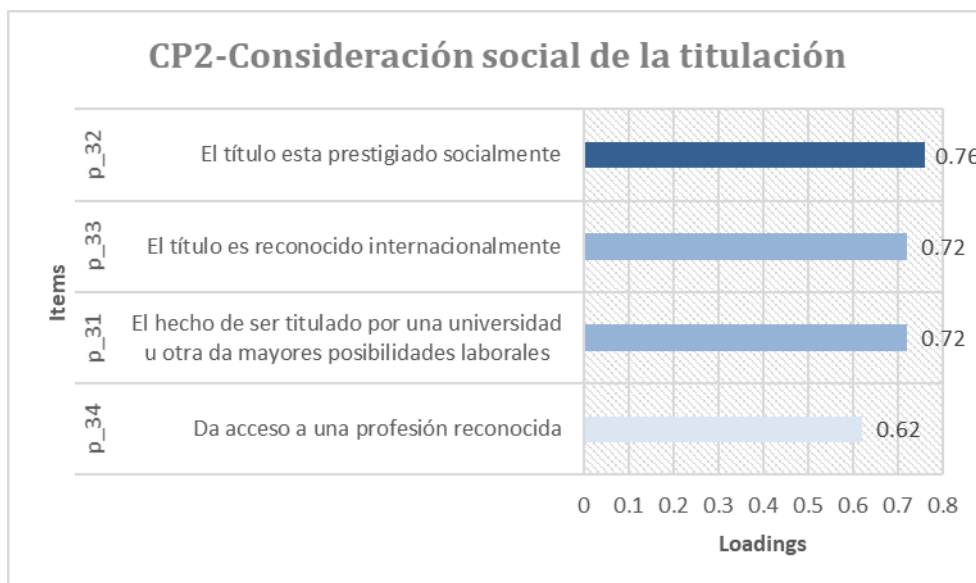


Gráfico 4. Segunda componente (PC2)

En la **tercera componente** (PC3) se observa dos variables con altas saturaciones, una variable positiva y otra negativa. La variable positiva hace alusión al deseo del estudiante de querer estudiar la carrera de su elección, mientras que la variable negativa indica que el estudiante eligió su carrera a última hora (ver Gráfico 5). Esta componente puede, por tanto, definirse como un componente que representa los **Aspectos Vocacionales** que llevaron al estudiante elegir un determinado grado o titulación. Por otro lado, esto nos indica que a mayor valoración de la variable “fue una decisión de última hora”, menor importancia tendrán los aspectos vocacionales.

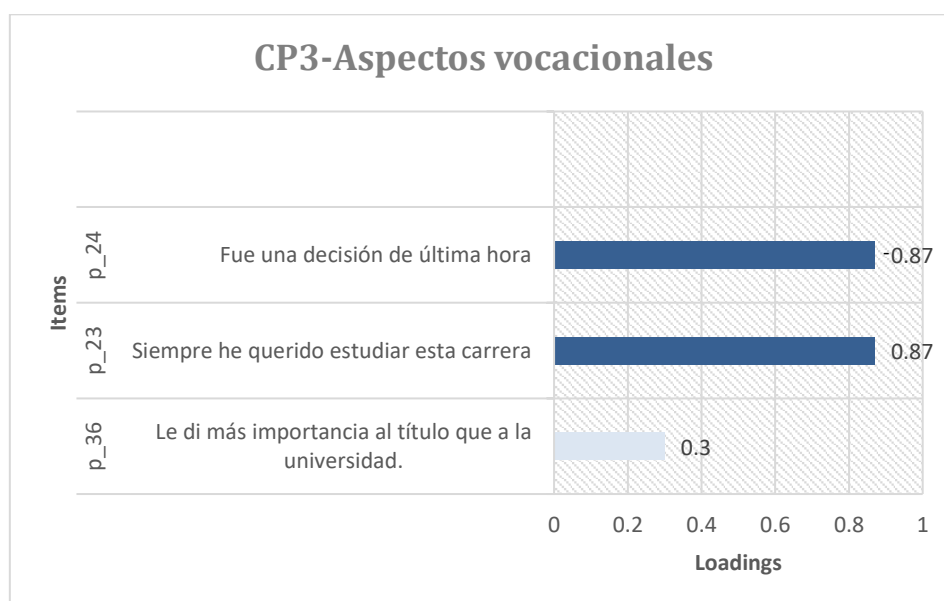


Gráfico 5. Tercera componente (PC3)

La **cuarta componente** principal (PC4) queda definida como **Entorno Próximo Académico** del estudiante, ya que esta componente relaciona la influencia decisiva de los profesores y orientadores en la elección de una carrera y en menor medida la influencia de los compañeros sobre dicha elección, los cuales se ven reflejados en las variables p_17 y p_16 con altas saturaciones apreciables en el Gráfico 6.

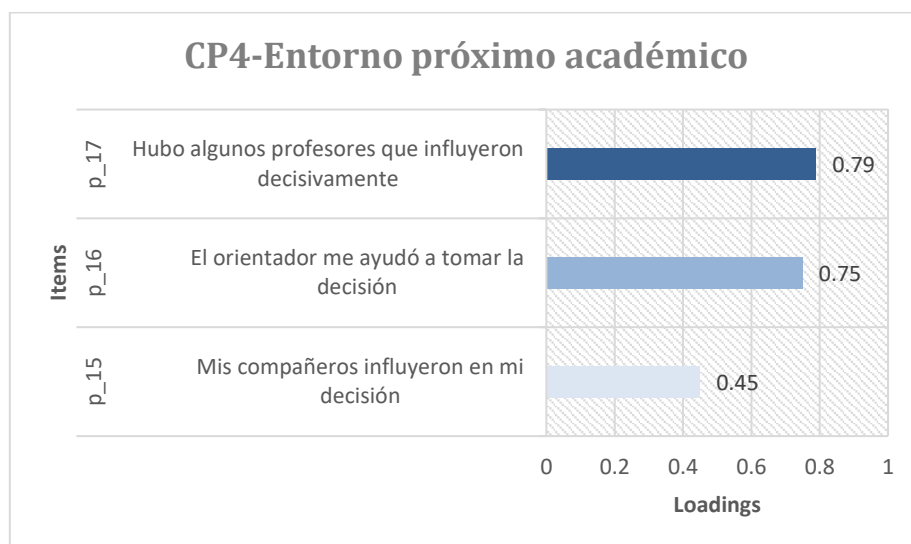


Gráfico 6. Cuarta componente (PC4)

En la **quinta componente** principal (PC5), las variables p_18 y p_19 guardan correlación entre sí (ver Gráfico 7). Estas dos variables están relacionadas con la información que proporcionan las universidades a los estudiantes mediante su página web y la visita al centro/universidad o la asistencia a las jornadas de acogida brindadas por la universidad. Por tal motivo, estas dos variables que conforman el quinto componente podría definirse como un componente que representa las diversas **Fuentes de Información** consultadas por los estudiantes durante el proceso de elección de determinado grado o titulación.

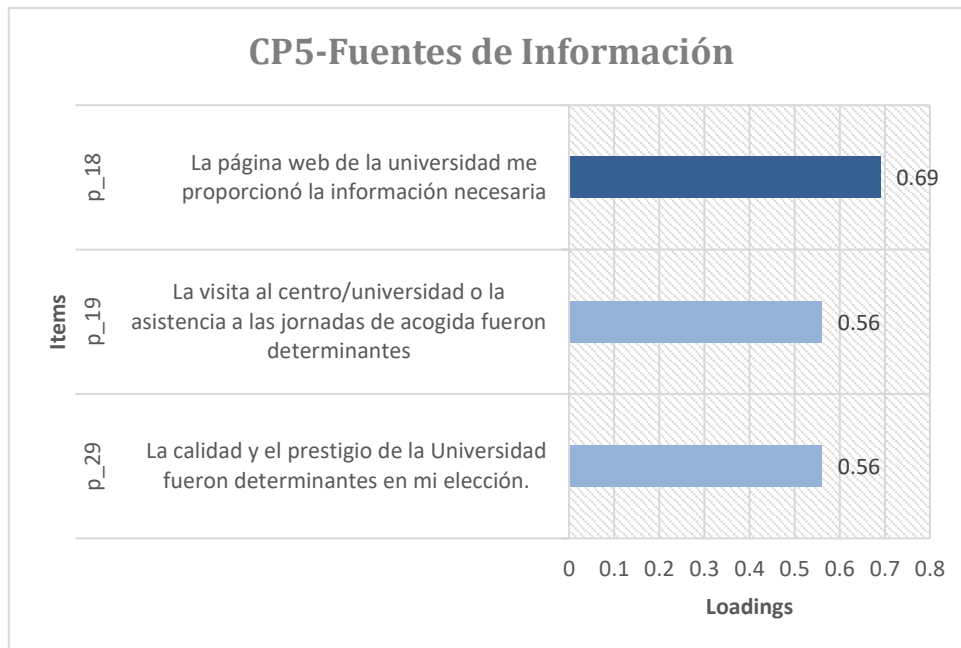


Gráfico 7. Quinta componente (PC5)

La **sexta componente** combina la obtención de buenas calificaciones en las asignaturas de bachillerato relacionados con la titulación elegida por el estudiante con la variable concerniente a las habilidades personales que el estudiante posee. Esto se debe a que tanto la variable p_26 como la variable p_25 tienen una alta saturación, mientras que la variable p_36 toma un valor muy bajo (ver Gráfico 8). La sexta componente principal (PC6) podría definirse como un componente que representa la **Adecuación Personal al Título o grado** que cursa el estudiante.

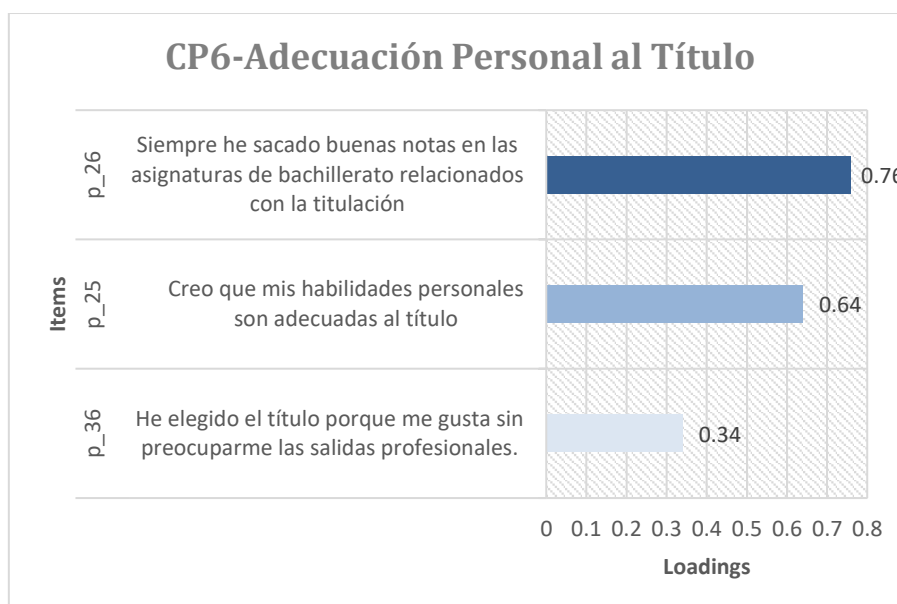


Gráfico 8. Sexta componente (PC6)

En la **séptima componente** principal (PC7), las variables p_20 y p_21 guardan correlación entre sí (ver Gráfico 9). Estas dos variables están relacionadas con la influencia que ejercen los padres y/o hermanos al haber estudiado una determinada carrera o haber estudiado en una determinada universidad frente a la elección personal del estudiante. Por tal motivo, estas dos variables conforman la séptima componente que podría definirse como la influencia del **Entorno Próximo Familiar** al que está sujeto el estudiante durante la elección de un grado universitario.

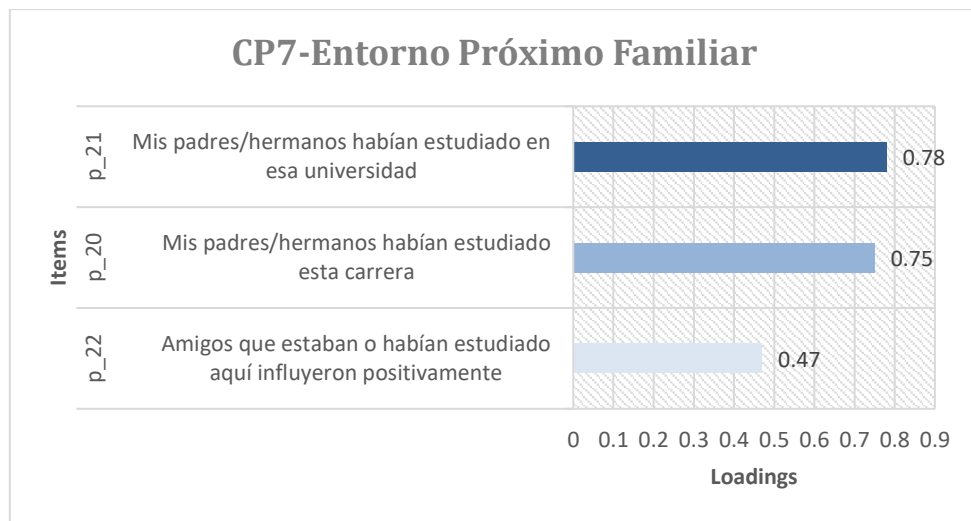


Gráfico 9. Sexta componente (PC7)

Por último, la **octava componente** principal (PC8) relaciona la proximidad del centro de estudios al domicilio del estudiante con la variable que determina que el estudiante le da más importancia al título que la universidad. Esto puede ser un indicativo de que el estudiante no le da importancia a la universidad en sí misma porque esta vendrá determinada por la cercanía a su domicilio y en cambio le dará más importancia al título. Esto se debe a que tanto la variable p_13 y la variable p_30 tienen una alta saturación, mientras que la variable p_12 toma un valor muy bajo (ver Gráfico 10). La octava componente principal podría definirse como un componente que representa la influencia de la **Proximidad Geográfica** en la elección del estudiante.

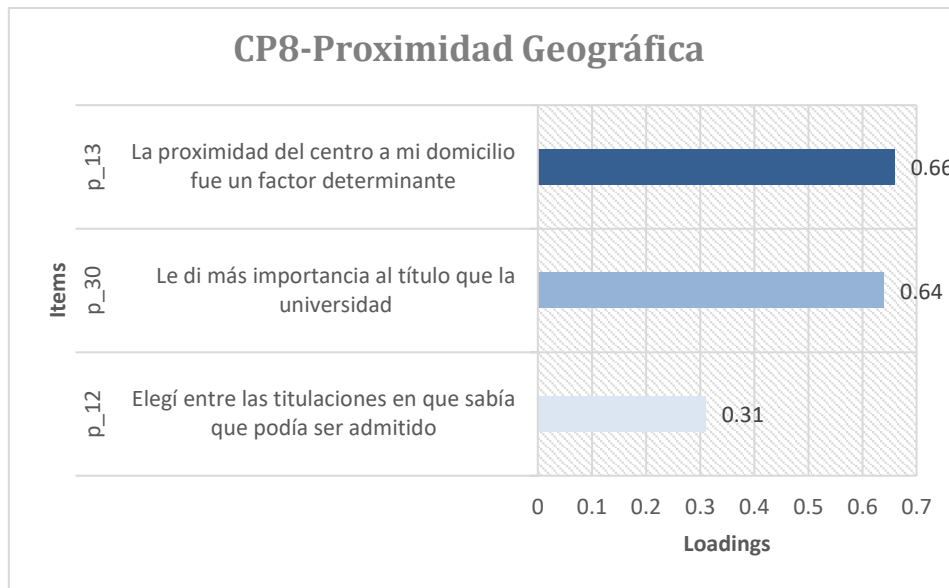


Gráfico 10. Octava componente (PC8)

En conclusión, analizando la matriz 1 de la Tabla 1 hemos podido designar un nombre a cada una de las ocho componentes principales extraídas, además de verificar la coherencia de los resultados obtenidos con la realidad interpretando las relaciones de cada uno de los componentes y los ítems.

A partir de los resultados obtenidos en este primer análisis pasaremos a realizar un análisis factorial exploratorio.

5.2 Análisis factorial

En la sección 4.2 del Capítulo IV mostramos los análisis preliminares realizados sobre los datos de la muestra. El resultado obtenido con el análisis del test de esfericidad de Bartlett fue ($p < .000$) y el del índice KMO fue 0.75. Por ello, llegamos a la conclusión de que nuestro conjunto de datos es factorizable y los resultados de los análisis preliminares garantizan que la aplicación de un AF ofrecerá resultados estables, lo que permitirá una interpretación adecuada de los factores.

Para la aplicación del proceso del Análisis Factorial Exploratorio (AFE) y Análisis Factorial Confirmatorio (AFC) se tendrá en cuenta la aproximación clásica, la cual recomienda hacer un uso secuencial de ambos tipos de análisis (Anderson y Gerbing, 1988; Brown, 2006). El primer paso consiste en dividir la muestra aleatoriamente en dos submuestras y extraer con la primera muestra, mediante la aplicación del AFE, la estructura factorial subyacente a los ítems. El segundo paso consiste en tratar de confirmar la estructura factorial resultante en la otra mitad de la muestra, esta vez mediante un AFC.

5.2.1 Análisis factorial exploratorio (AFE)

La aplicación de AFE se subdivide en dos fases: la extracción del número de factores y la interpretación de los mismos.

5.2.1.1 Número de factores

Para determinar el número de factores a extraerse se utilizó el método de análisis paralelo, el cual se muestra en el Gráfico 11, donde el eje “X” posiciona a los factores y el eje “Y” corresponde a los autovalores.

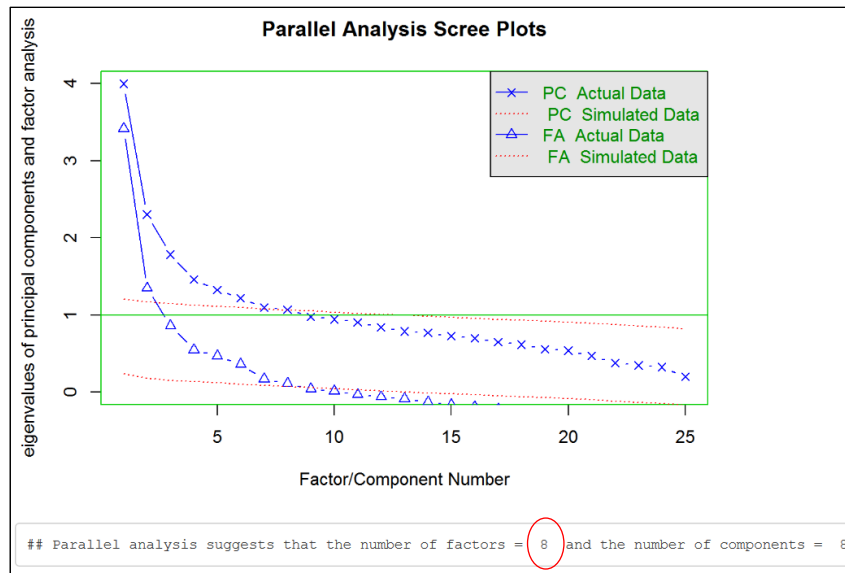


Gráfico 11. Análisis Paralelo

El Gráfico 11 muestra el resultado del análisis paralelo, el cual genera resultados tanto para el ACP como para AF y donde se puede observar que los resultados en ambos análisis son coincidentes en el número de componentes/factores a extraer. El número de factores a extraer se identifica mediante los autovalores generados por los datos originales, identificados en la leyenda del Gráfico 11 como (FA Actual Data), que quedan por encima de la curva generada por los autovalores de los datos aleatorios, identificados en la leyenda del Gráfico 11 como (FA Simulated Data). El número de autovalores que quedan por encima de dicha curva son ocho, que a su vez coincide con los datos del PCA identificados en la leyenda del Gráfico 11 como (PC Actual Data y PC Simulated Data).

Una vez identificado el número de factores a extraer, se procede a generar la matriz factorial aplicando para ello el método de la máxima verosimilitud (MLE - Maximum Likelihood Estimation), uno de los métodos de estimación clásica de parámetros más empleada en el análisis factorial.

Tras la aplicación del método MLE se obtuvo la matriz factorial (sin rotación). En dicha matriz no se observaron patrones significativos en las cargas factoriales de las variables. Por dicho motivo, se procedió a aplicar la rotación oblicua (Oblimin) a la matriz y que este método es el más congruente con la estructura de las variables no observadas directamente que, por lo general, se entienden como pertenecientes a un mismo factor.

El resultado de aplicar la rotación oblicua hace que los pesos factoriales de una variable sean lo más cercana a 1 con uno solo de los factores, estableciendo cargas factoriales próximas a 0 con el resto de factores. En otras palabras, permite obtener una matriz factorial más simple y fácil de interpretar.

En la Tabla 2 se muestran dos matrices. La primera se corresponde con la matriz rotada resultante del AFE (Matriz 1). En la primera fila de la matriz factorial se aprecia el nombre de cada factor (F1, F2, ..., F8), mientras que en la primera columna se encuentran las 25 variables codificadas. La intersección entre un elemento de la columna y uno de la fila representa el valor del peso factorial que el ítem correspondiente aporta a cada factor (se ha suprimido las cargas factoriales inferiores a 0.30 para facilitar su lectura).

##	F1	F2	F3	F4	F5	F6	F7	F8	h2	u2	com
## p_12									0.105	0.895	1.6
## p_13									0.084	0.916	4.2
## p_14									0.036	0.964	2.9
## p_15									0.112	0.888	2.6
## p_29				0.89					0.830	0.170	1.0
## p_16					0.53				0.337	0.663	1.2
## p_30									0.109	0.891	2.6
## p_17					0.77				0.595	0.405	1.0
## p_31	0.33			0.31					0.222	0.778	2.6
## p_18								0.46	0.263	0.737	1.3
## p_32	0.77								0.593	0.407	1.0
## p_19								0.51	0.351	0.649	1.2
## p_33	0.78								0.597	0.403	1.0
## p_20							0.51		0.275	0.725	1.2
## p_34	0.74								0.687	0.313	1.1
## p_21							0.63		0.396	0.604	1.1
## p_22									0.183	0.817	2.8
## p_36									0.230	0.770	3.0
## p_23				1.00					0.995	0.005	1.0
## p_37	0.30	0.43							0.451	0.549	2.1
## p_24				-0.58					0.391	0.609	1.1
## p_38		0.84							0.711	0.289	1.0
## p_25						0.44			0.292	0.708	1.4
## p_39		0.93							0.861	0.139	1.0
## p_26						0.52			0.315	0.685	1.2
##											
##				F1	F2	F3	F4	F5	F6	F7	F8
## SS loadings				2.14	1.94	1.49	1.09	1.00	0.88	0.80	0.69
## Proportion Var				0.09	0.08	0.06	0.04	0.04	0.04	0.03	0.03
## Cumulative Var				0.09	0.16	0.22	0.27	0.31	0.34	0.37	0.40
## Proportion Explained				0.21	0.19	0.15	0.11	0.10	0.09	0.08	0.07
## Cumulative Proportion				0.21	0.41	0.56	0.66	0.76	0.85	0.93	1.00

Matriz 1

Matriz 2

Tabla 2. Resultados de AFE con rotación oblicua

Para este caso, al igual que en la matriz resultante de ACP, se ha resaltado mediante círculos de colores las cargas factoriales más altas de las variables que contribuyen a cada factor. Así se observa que en el factor 1 (F1), las variables que tienen las cargas factoriales más altas son p_32, p_33 y p_34, lo que implica que este factor queda definido con estas tres variables. Las variables p_38 y p_39, se aprecia que tienen cargas factoriales altas en el segundo factor (F2), lo que hace que el segundo factor quede definido mediante estas dos variables.

De igual forma, así sucesivamente podemos ir observando cómo en base a las cargas factoriales más altas de las variables se definen cada uno de los factores.

Además de las cargas factoriales, en la Matriz 1 se muestra el valor de la Comunalidad (h^2), que es la proporción de la varianza explicada por los factores comunes en una variable, y la Unicidad (u^2), que es el porcentaje de la varianza que no ha sido explicada por el factor y se calcula como $1 - \text{Comunalidad}$.

Los valores de la Matriz 2 de la Tabla 2 son los siguientes:

- La fila 'SS Loadings' muestra la medida de la varianza que explica cada factor. En este caso solo cinco factores superan a la unidad, lo que nos indica que, si hubiéramos utilizado la regla de Kaiser para seleccionar el número de factores a retener, solo cinco serían los factores. Sin embargo, esto solo representaría un 31% de la proporción de la varianza, un valor poco recomendado para este tipo de análisis.
- En la fila 'Proportion Var' se puede apreciar la proporción de la varianza que aporta cada factor al total de la varianza acumulada de los ocho factores.
- En la fila 'Cumulative Var' indica que la extracción de ocho factores representa el 40% de la varianza total de los datos, un 16% menos que el obtenido en el ACP. Esto puede justificarse debido a que se está trabajando con una submuestra aleatoria, la cual representa la mitad de los datos de la muestra total. Además, el análisis factorial trabaja con la varianza común a todas las variables y esta varianza es la que indica la diferencia que se crea en las respuestas.

A continuación, en base a los resultados de la Tabla 2 (Matriz 1) se procedió a realizar la interpretación de cada uno de los ocho factores extraídos.

5.2.1.2 Interpretación de factores

A partir de los resultados de la Tabla 2 se generó un diagrama donde se puede observar con mayor claridad cada uno de los factores y sus respectivos ítems (ver Gráfico 12).

El **primer factor** (F1) es claramente el factor de la **Consideración social del título** que cursa el estudiante, ya que las cargas factoriales más altas están presentes en las variables relacionadas con la percepción social vinculado con el prestigio o la relevancia y nivel de internacionalización del título, mientras que la carga factorial más baja hace referencia a la percepción de que el ser titulado por una determinada universidad genera mayores posibilidades laborales. Por tal motivo, estas cuatro variables (p_32, p_33, p_34 y p_31) conforman el primer factor (ver Tabla 2).



El **segundo factor** (F2) combina la facilidad de encontrar trabajo y la creencia de que existe una demanda en el sector, y en menor medida la creencia de que los salarios que se consiguen con cierta titulación son mejores que con otras titulaciones. Esto se debe a que las variables p_39 p_38 y p_37 muestran cargas factoriales altas los cuales se pueden apreciar en la Tabla 2. El segundo factor podría definirse como un factor que representa la **Empleabilidad percibida de la titulación o grado** que cursa el estudiante.

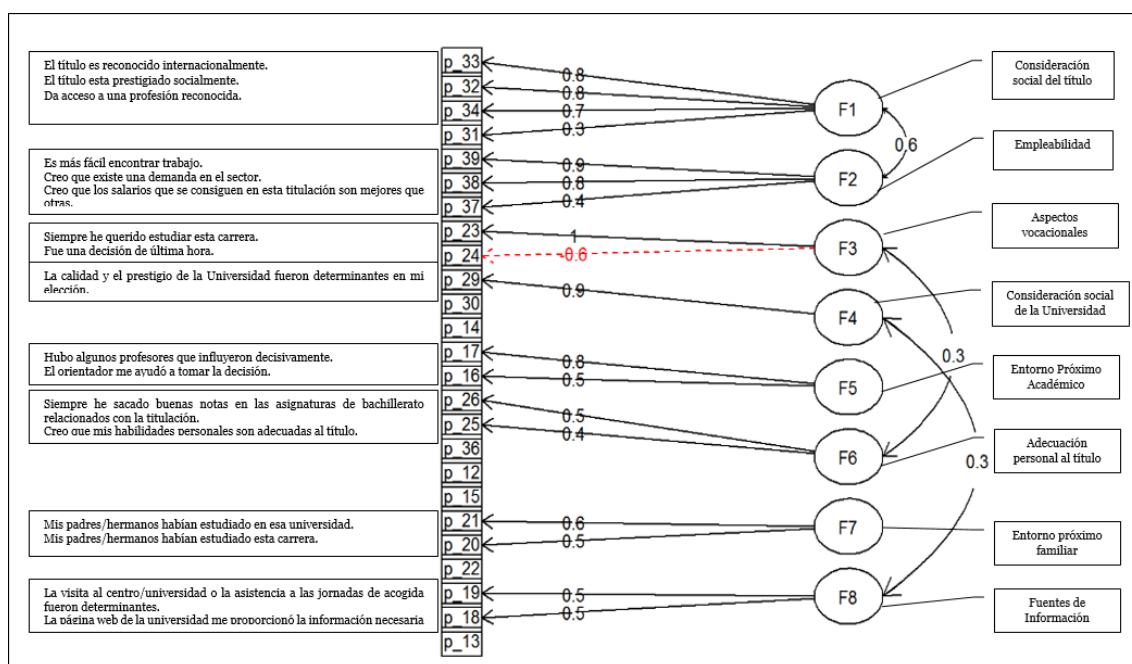


Gráfico 12. Diagrama de AFE

El **tercer factor** (F3) muestra cargas factoriales altas en dos variables. La variable p_23 presenta una carga factorial positiva, mientras que la segunda variable p_24 posee una carga factorial negativa (ver Tabla 2). El hecho de que la carga de la variable p_24 sea negativo significa que se calificó como el extremo opuesto de la variable p_23. Este tercer factor hace alusión a que el estudiante siempre ha deseado estudiar la carrera de su elección, por tanto, no se trata de una decisión de última hora. Por tal motivo, este factor podría definirse como un factor que representa los **Aspectos Vocacionales** que llevaron al estudiante elegir un grado o titulación

El **cuarto factor** (F4) queda definido por una única variable p_29 que presenta una carga factorial alta, casi cercana a uno (ver Tabla 2). Este factor podría definirse como un elemento que representa la **Consideración social de la universidad**, ya que hace referencia a que la calidad y el prestigio de la Universidad fueron determinantes en la elección de un grado universitario por parte del estudiante. Cabe resaltar que este factor no aparece en los resultados obtenidos del ACP.

El **quinto factor** (F5) relaciona la influencia decisiva de los profesores y orientadores en la elección de una carrera. Esto se debe a que tanto la variable p_16 como la variable p_17 muestran altas cargas factoriales (ver Tabla 2). Este quinto factor podría definirse como un factor que representa la influencia que ejerce el **Entorno Próximo Académico** en la elección de un grado o titulación.

El **sexto factor** (F6) combina la obtención de buenas calificaciones en las asignaturas de bachillerato relacionados con la titulación elegida por el estudiante con la variable concerniente a las habilidades personales que el estudiante posee. Esto se debe a que tanto la variable p_26 como la variable p_25 tienen altas cargas factoriales (ver Tabla 2). Este sexto factor podría definirse como un factor que representa la **Adecuación Personal al Título o grado** que cursa el estudiante.

El **séptimo factor** (F7) relaciona la influencia que ejercen los padres y/o hermanos al haber estudiado una determinada carrera o haber estudiado en una determinada universidad frente a la elección personal del estudiante. Esto se debe a que las dos variables, p_20 y p_21 tienen altas cargas factoriales (ver Tabla 2). Por tal motivo, estas dos variables que conforman el séptimo factor podría definirse como un factor que representa la influencia del **Entorno Próximo Familiar** al que está sujeto el estudiante durante la elección de un grado universitario.

Por último, en el **octavo factor** (F8) son dos las variables (p_19 y p_18) que presentaron las más altas cargas factoriales (ver Tabla 2). Estas dos variables combinan la información que proporcionan las universidades a los estudiantes mediante su página web y la visita al centro/universidad o la asistencia a las jornadas de acogida brindadas por la universidad. Este sexto factor podría definirse como un factor que representa la **Adecuación Personal al Título o grado** que cursa el estudiante.

Finalmente, las diferencias más significativas entre los 8 componentes obtenidos del PCA y los 8 factores extraídos en el AFE se localiza en el factor denominado *Consideración social de la universidad*, y en el factor *Proximidad geográfica* que se ha encontrado en el ACP, pero no en el AFE. Esto podría justificarse porque el AFE analiza la varianza común a todas las variables, partiendo de una matriz de correlaciones o covarianzas, mientras que el ACP básicamente trata de explicar la mayor proporción posible de la varianza total de las variables. Por otro lado, cabe mencionar que la muestra con la se está trabajando el AFE no es la misma que en el ACP. En el AFE se trabaja con la mitad de la muestra total. Y, aun así, se puede comprobar que la estructura de componentes/factores es muy similar en ambos análisis.

5.2.2 Análisis factorial confirmatorio (AFC)

A partir del diagrama resultante del AFE (Gráfico 12), se procedió a validar el modelo resultante con la otra mitad de los datos, tal y como se tenía previsto en la formulación de la aplicación del AF. La primera observación es que la validación del modelo que se muestra en el Gráfico 12 no presentaba convergencia. Esto significa que no se obtuvieron resultados para los indicadores de ajuste de validación del modelo descritos en la sección 3.3.5. Por ello, se procedió a modificar el modelo generado por AFE partiendo del

supuesto de que el ítem que genera la no convergencia se encuentra en uno de los ítems con el peso factorial más bajo dentro de cada factor. La idea es identificar y excluir el ítem que causa la no convergencia del modelo mediante los siguientes pasos:

Paso 1: para la construcción del que llamaremos el primer modelo se realizaron varias pruebas previas eliminando un solo ítem en cada prueba.

- Primero se identificaron los ítems con los pesos factoriales más bajos para cada factor, entre los cuales se identificaron los factores F1, F2, F3, F5, F6, F7 y F8, y los ítems (p_31, p_37, p_24, p_16, p_25, p_20 y p_18) respectivamente (ver Gráfico 12).
- Posteriormente, en esta primera prueba, solo se excluyó el primer ítem del primer factor (p_31) conservando los demás ítems restantes. Para la segunda prueba solo se excluyó el ítem del segundo factor (p_37), igualmente conservando los demás ítems restantes incluido el ítem p_31, y así sucesivamente. Es decir, se realizó un total de 7 ajustes previos, quitando uno a uno el ítem identificado como el peso más bajo en cada factor hasta obtener un modelo con resultados para cada uno de los indicadores de ajuste propuestos.
- Finalmente, tras realizar los ajustes descritos se encontró una primera solución óptima, la cual excluía únicamente el ítem p_20 que tiene asociado un peso de 0.5, pero conservando los restantes ítems (ver Tabla 3). Cabe resaltar que este ítem tiene el peso factorial más bajo dentro del séptimo factor relacionado con la influencia del entorno próximo familiar.

Modelo	Factores	ítems
M1	Consideración social de la titulación	p_33, p_32, p_34, p_31
	Empleabilidad	p_39, p_38, p_37
	Aspectos vocacionales	p_23, p_24
	Consideración social de la Universidad	p_29
	Entorno próximo académico	p_17, p_16
	Adecuación personal al título	p_26, p_25
	Entorno Próximo Familiar	p_21
	Fuentes de Información	p_19, p_18

Tabla 3. Definición del primer modelo

Las estimaciones de los índices propuestos para la validación resultantes del modelo (M1) se presentan en Tabla 4.

chi-cuadrado (χ^2)	Grados de libertad (gl)	(χ^2) / gl	CFI	TLI	RMSEA	SRMR
813.854	93	8.751118	0.917	0.878	0.06	0.038

Tabla 4. Estadísticos de ajuste del (M1)

Para medir la calidad del ajuste de cada uno de los modelos se emplearon los seis indicadores más utilizados en este tipo de análisis, los cuales se han descrito en detalle en la sección 3.3.5. A continuación, se presenta un resumen de los valores de referencia de estos seis indicadores.

El primer indicador conocido como la prueba Chi- cuadrado (X^2) sugiere que el valor resultante debe ser dos o tres veces mayor que los grados de libertad para ser aceptable. El segundo indicador es la razón ($X^2/\text{grados de libertad}$), el cual se considera un buen indicador si el resultado oscila entre uno y tres. El tercer indicador, conocido como el índice de Ajuste Comparativo (CFI, por sus siglas en inglés, Comparative Fit Index), compara la mejora en el ajuste del modelo en cuestión con un modelo alternativo. El cuarto indicador, llamado índice de Tucker-Lewis (TLI, por sus siglas en inglés, Tucker-Lewis Index), refleja la proporción de información total explicada por un modelo. El valor de los indicadores CFI y TLI deben ser cercanos a 1 para que indique un muy buen ajuste al modelo. Entre los índices basados en las covarianzas se encuentra el quinto indicador, el índice de Aproximación de la Raíz de Cuadrados Medios del Error (RMSEA, por sus siglas en inglés, Root Mean Square Error of Aproximation), valores menores a 0.06 indican un buen ajuste. Por último, para el sexto indicador llamado índice de la Raíz de los Cuadrados Medios del Residuo (SRMR, por sus siglas en inglés, Standardized Root Mean squared Residuals), valores menores a 0.08 indican un buen ajuste.

En la Tabla 4 observamos que el valor de la prueba X^2 no aporta un valor aceptable, ya que es un valor mucho mayor que 3 veces los grados de libertad. La razón ($X^2/\text{grados de libertad}$) es mayor a 3, lo que tampoco representa un valor adecuado. Los valores de CFI y TLI se podrían considerar como adecuados, ya que sus valores son cercanos a 1. Por último, de los valores de RMSEA y SRMR, solo el segundo tiene un resultado óptimo (menor a 0.05).

Debido a que los resultados obtenidos de los indicadores del primer modelo (Modelo M1 de la Tabla 3) no son adecuados de acuerdo a los valores de referencia, se procede a ajustar el primer modelo a fin de continuar la validación.

Paso 2: el segundo modelo se generó en base al primer modelo eliminando los pesos factoriales menos representativos, aquellos con un valor menor a 0.5.

- Del factor relacionado con la consideración social de la titulación se excluye el ítem p_31, el cual tenía un valor de 0.33.
- Del factor relacionado con la empleabilidad se excluye el ítem p_37, con un valor de 0.43.
- Del factor relacionado con la adecuación personal al título se excluye el ítem p_25, con un valor de 0.44.
- Del factor relativo a las fuentes de información no se considera el ítem p_18, con un valor de 0.46.

Tras estos ajustes, el nuevo modelo resultante M2 es el que se muestra en la Tabla 5.

Modelo	Factores	ítems
M2	Consideración social de la titulación	p_33 , p_32, p_34
	Empleabilidad	p_39, p_38
	Aspectos vocacionales	p_23, p_24
	Consideracion social de la Universidad	p_29
	Entorno próximo académico	p_17, p_16
	Adecuación personal al título	p_26
	Entorno Próximo Familiar	p_21
	Fuentes de Información	p_19

Tabla 5. Definición del segundo modelo

Las estimaciones de los índices propuestos para la validación resultantes del modelo (M2) se presentan en Tabla 6.

chi-cuadrado (χ^2)	Grados de libertad (g1)	(χ^2) / g1	CFI	TLI	RMSEA	SRMR
176.656	41	4.308683	0.978	0.959	0.039	0.019

Tabla 6. Estadísticos de ajuste del (M2)

La primera observación que podemos deducir de la Tabla 6 es que los valores de todos los indicadores de ajuste han mejorado respecto a los de la Tabla 4. No obstante, si analizamos el valor de la prueba X^2 podemos observar que, al igual que el primer modelo M1, tampoco aporta un valor aceptable, ya que es un valor superior a tres veces los grados de libertad. La razón (X^2 /grados de libertad) si bien es cierto que existe una gran reducción respecto al valor obtenido en el primer modelo, aun es mayor a 3, lo que tampoco representa un valor adecuado. Los valores de CFI y TLI son óptimos, muy cercanos a 1. Por último, los valores de RMSEA y SRMR presentan una mejora sustancial respecto al primer modelo, ambos presentan un resultado óptimo (menor a 0.05).

Si bien es cierto que los ajustes realizados en el segundo modelo presentan valores más óptimos en los cuatro últimos índices, aún hay dos índices que no se acercan a los valores de referencia dados, motivo por el cual se procede a realizar un tercer ajuste al modelo.

Paso 3: para el diseño del tercer modelo se decide conservar hasta un máximo de dos ítems para cada factor teniendo en cuenta que los pesos factoriales fueran los más representativos; el resultado del tercer modelo M3 se puede apreciar en la Tabla 7.

Modelo	Factores	ítems
M3	Consideración social de la titulación	p_33, p_32
	Empleabilidad	p_39, p_38
	Aspectos vocacionales	p_23, p_24
	Consideración social de la Universidad	p_29
	Entorno próximo académico	p_17, p_16
	Adecuación personal al título	p_26
	Entorno Próximo Familiar	p_21
	Fuentes de Información	p_19

Tabla 7. Definición del tercer modelo

Las estimaciones de los índices propuestos para la validación resultantes del modelo (M3) se presentan en Tabla 8.

chi-cuadrado (χ^2)	Grados de libertad (gl)	(χ^2)/gl	CFI	TLI	RMSEA	SRMR
89.233	30	2.974433	0.987	0.972	0.03	0.015

Tabla 8. Estadísticos de ajuste del (M3)

En la Tabla 8 podemos apreciar que el tercer modelo M3 presenta los mejores ajustes de calidad, ya que el valor de la prueba X^2 se encuentra entre 2 y 3 veces el valor de los grados de libertad, la razón (X^2 /grados de libertad) es menor a 3, los valores de CFI y TLI son mayores a 0.95, y valores de RMSEA y SRMR menores a 0.05 indican buen ajuste del modelo.

Un resumen de las estimaciones de los indicadores de ajuste para los tres modelos de AFC se presentan en Tabla 9.

Modelo	chi-cuadrado (χ^2)	Grados de libertad (gl)	(χ^2)/gl	CFI	TLI	RMSEA	SRMR
M1	813.854	93	8.751118	0.917	0.878	0.06	0.038
M2	176.656	41	4.308683	0.978	0.959	0.039	0.019
M3	89.233	30	2.974433	0.987	0.972	0.03	0.015

Tabla 9. Estadísticos de ajuste para los modelos de AFC

De la tabla 9 podemos observar que los indicadores de ajuste del tercer modelo M3 son óptimos (se cumplen los valores de referencia de los seis indicadores propuestos) en comparación con los dos primeros modelos; es decir, el modelo ajusta adecuadamente los datos. Cabe resaltar que la etapa en la que se lleva a cabo la evaluación de los modelos

propuestos es fundamental para determinar si dicho modelo describe de manera apropiada al fenómeno objeto de estudio.

La mayoría de los factores del tercer modelo M3 (ver Tabla 7) quedó compuesta por dos ítems cada uno, salvo los factores referentes a la *Consideración social de la universidad*, la *Adecuación personal del título*, el *Entorno próximo familiar* y las *Fuentes de información* que desde el inicio tenían ítems poco representativos que no aportaban a la obtención de un buen ajuste al modelo.

Finalmente, teniendo en cuenta cada uno de los pasos seguidos en el AFC, se observa que difícilmente los modelos de AFC ajustan los datos en un primer análisis, pero el diagnóstico no sólo ha permitido evaluar el modelo, sino que también ha ayudado a mejorarlo. Así, la modificación del modelo inicial generado por el AFE ayuda a optimizar la bondad del ajuste mediante la supresión de ciertos ítems. Cabe resaltar que también se realizaron algunas pruebas alternativas mediante la generación de varias submuestras provenientes aleatoriamente de la muestra original. De los resultados obtenidos todos ellos también sugerían el retiro de la variable p_20 a fin de obtener un resultado inicial satisfactorio.

Conclusiones

A lo largo del desarrollo del presente trabajo el objetivo principal ha sido estudiar e identificar los factores principales que motivan a los estudiantes en la elección de un grado universitario mediante la aplicación de dos técnicas de análisis estadístico, ACP y AF, sobre una muestra de datos de casi 4,400 estudiantes a un cuestionario de 25 preguntas. Un análisis previo a la aplicación del ACP y AF revela que el conjunto de datos de la muestra es factorizable y, por tanto, es posible obtener resultados estadísticamente significativos.

La aplicación de estas dos técnicas de análisis estadístico, ACP y AF, devuelve un mismo número de factores latentes, lo que significa que es posible explicar las motivaciones que influyen en los estudiantes en la elección de un grado universitario mediante ocho factores latentes.

A partir del análisis de las 25 variables del cuestionario con el ACP se encontró una solución con ocho componentes principales: el primer componente se define claramente como la **Empleabilidad**, el segundo como la **Consideración Social del título**, el tercero los **Aspectos Vocacionales**, el cuarto relacionado con el **Entorno Próximo Académico**, el quinto se define como las **Fuentes de Información**, el sexto como la **Adecuación Personal al Título**, el séptimo el **Entorno Próximo Familiar** y, por último, el octavo, definida como la **Proximidad Geográfica**.

Por otra parte, la estructura de factores encontrada en el AFE indica que los factores relacionados con la elección de un grado universitario pueden reducirse a ocho factores, el primer factor encontrado es la **Consideración social del título**, el segundo la **Empleabilidad**, el tercero los **Aspectos vocacionales**, el cuarto la **Consideración social de la universidad**, el quinto el **Entorno próximo académico**, el sexto la **Adecuación personal al título**, el séptimo como el **Entorno próximo familiar** y, el octavo como las **Fuentes de información**.

Se puede evidenciar a partir de los análisis existen unos claros componentes o factores que intervienen en la elección de un grado universitario.

De los resultados obtenidos con el ACP y el AF se puede observar que existe una diferencia en cuanto a la interpretación del nombre de los factores o componentes principales. En el caso del AFE se observa que aparece un factor que hace alusión a la consideración social de la universidad, en cambio, no se encuentra al componente relacionado con la proximidad geográfica. La justificación puede deberse a que, a pesar de que tanto el ACP como el AFE tienen como objetivo principal simplificar la información de un gran conjunto de datos, el ACP lo hace analizando la varianza total en las variables, mientras que el AFE lo hace a través del análisis de la varianza común a todas las variables. Además, para el cálculo del ACP se utilizó la muestra total de los datos, mientras que para el AFE se utilizó una submuestra aleatoria, la cual representa la mitad de los datos de la muestra total.

A pesar de alguna discrepancia en la interpretación de los factores, se puede constatar que ambas técnicas identifican unos factores comunes.

Por otro lado, la consideración o prestigio del título y la empleabilidad son los dos factores más determinantes detectados por ambas técnicas de análisis de datos.

Por tanto, los resultados de ambos análisis conducen a la conclusión de que se puede realizar un test que contemple una o varias preguntas relacionadas con cada uno de los factores identificados.

A partir de la estructura de factores generada por el AFE se obtuvo un modelo que fue validado a través de la aplicación del AFC. Dicho modelo posee una aceptable validez dado que los valores de todos los índices de ajuste (CFI, TLI > .95 y RMSEA, SRMR < .05), se encuentran dentro de los límites de referencia, indicando así que la nueva distribución y supresión de ítems sugerida en el AFC es adecuada para sugerir un nuevo modelo de encuesta.

En conclusión, se puede construir un cuestionario con la mitad de preguntas del cuestionario original y se obtendría la misma información porque dichas preguntas recogerían la misma variabilidad. Esto es, la elección de un grado universitario se podría analizar por medio de ocho factores conformados por doce ítems distribuidos de la siguiente manera:

- **Consideración social del título**
 - El título está prestigiado socialmente.
 - El título es reconocido internacionalmente.
- **Empleabilidad**
 - Es más fácil encontrar trabajo.
 - Creo que existe una demanda en el sector.
- **Aspectos vocacionales**
 - Siempre he querido estudiar esta carrera.
 - Fue una decisión de última hora.
- **Consideración social de la universidad**
 - La calidad y el prestigio de la Universidad fueron determinantes en mi elección.
- **Entorno próximo académico**
 - El orientador me ayudó a tomar la decisión.
 - Hubo algunos profesores que influyeron decisivamente.
- **Adecuación personal al título**
 - Siempre he sacado buenas notas en las asignaturas de bachillerato relacionados con la titulación.
- **Entorno próximo familiar**
 - Mis padres/hermanos habían estudiado en esa universidad.
- **Fuentes de información**
 - La visita al centro/universidad o la asistencia a las jornadas de acogida fueron determinantes.

Referencias

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Computational Statistics*, 2(4), 433-459.
- Anderson, J. C., y Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.
- Blalock, H. M. (1966). *Estadística social*. México D. F., México: Fondo de Cultura Económica.
- Bisquerra, R. (1989). *Introducción conceptual al análisis multivariable. Un enfoque informático con los paquetes SPSS-X, BMDP, LISREL y SPAD*. Barcelona, España: PPU.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Cabrera, L., Bethencourt, J. T., Pérez, P. A., & Afonso, M. G. (2006). *El problema del abandono de los estudios universitarios*. Recuperado de <http://www.redalyc.org/pdf/916/91612201.pdf>
- Carmona, F. (2014). *Un ejemplo de ACP paso a paso. Apuntes*. Barcelona, España: Universidad de Barcelona.
- Cattell, R. (1966). The Scree Test for the number of factors. *Multivariate Behavioral Research*. 1, 141-161.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(7), 1-9.
- Cuadras, M. C., (2008). *Nuevos métodos de análisis multivariante*. Barcelona, España: CMC.
- De la Fuente Fernández, S. (2011b). Análisis Componentes Principales-ACP, Material de apoyo docente, Santiago de la Fuente Fernández, Fac. Económica y de Ciencias Empresariales, Univ. Autónoma de Madrid (UAM).
- Fadlelmula, F. K. (2011). Assessing power of structural equation modeling studies: A meta-analysis. *Education Research Journal*, 1(3), 37-42.
- Fernández, C., Peña, V., Viñuela, P., & Torío, S. (2007). Los procesos de orientación escolar y la toma de decisiones académica y profesional. *Revista Complutense de Educación*, 18(2), 87-103
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: Algunas consideraciones adicionales. *Anales de Psicología*, 30(3), 1170-1175.
- Elias-Andreu, M. (2008). Los abandonos universitarios: Retos ante el espacio europeo de educación superior. *Estudios sobre Educación*, 15, 101-121



- Frías-Navarro, D., & Soler, M. P. (2013). *Prácticas del análisis factorial exploratorio (AFE) en la investigación sobre conducta del consumidor y marketing*. Recuperado de www.elsevier.es/sumapsicol.
- Fundación BBVA. (2010). *Estudio internacional Fundación BBVA sobre estudiantes universitarios de seis países europeos*. Recuperado de https://w3.grupobbva.com/TLFU/dat/resultados_universitarios_2010.pdf
- Gelin, M. N., Beasley, T. M., & Zumbo, B. D. (2003, April). What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale? Paper presented at the annual meeting of the American Educational Research Association (AERA) in Chicago, Illinois.
- González M., Díaz de Pascual, P. A., Torres, E., & Garnica, E. (1994). Una aplicación del análisis de componentes principales en el área educativa. *Economía*, 9, 55-72.
- Gorsuch, R. L. (1983). *Factor analysis*. (2nd. Ed.). Hillsdale, NJ: Erlbaum.
- Hervás, A., Guàrdia Olmos, J., Però Cebollero, M., Capilla Lladró, R., & Soriano Jiménez, P. P. (2013, August). A structural equation model for analysis of factors associated with the choice of engineering degrees in a technical university. *In Abstract and Applied Analysis* (Vol. 2013). Hindawi Publishing Corporation.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Jiménez, A., & Royo, A. (1993). Investigación sobre las variables relacionadas con la elección de estudios superiores. *Revista de Educación de la Universidad de Granada*, 7, 187-224.
- Jolliffe, I. T. (2002). *Principal component analysis*. (2nd ed). New York, NY: Springer.
- Kahn, J. H. (2006). Factor analysis in counseling psychology research, training and practice. *The Counseling Psychologist*, 34, 1-36.
- Kerlinger, F. N., & Lee, H. B. (2001). Investigación del comportamiento: métodos de investigación en ciencias sociales. *Mexico: McGraw-Hill*.
- Kline, P. (2014). *An easy guide to factor analysis*. London, United Kingdom: Routledge.
- Lagares, A. D. R. C., Lagares, E. C., & Ordaz, F. G. (2012). Empirical evidence on the process of choice of a university degree/Preuves empiriques sur le processus de choix d'un diplôme universitaire. *Enseñanza & Teaching*, 30(2), 45.
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological bulletin*, 97(3), 562.

- Martín, P. G., de Pascual, A. D., Lezama, E. T., & Olmos, E. G. (1994). Una aplicación del análisis de componentes principales en el área educativa. *Economía*, 19(9), 55-72.
- McIver, J., & Carmines, E. G. (1981). *Unidimensional scaling* (Vol. 24). Sage.
- Merenda, P. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. *Measurement and Evaluation in Counseling and Evaluation*, 30, 156-163.
- Morata-Ramírez, M., Holgado-Tello, F. P., Barbero-García, I., & Mendez, G. (2015). Análisis factorial confirmatorio: recomendaciones sobre mínimos cuadrados no ponderados en función del error Tipo I de Ji-Cuadrado y RMSEA. *Acción Psicológica*, 12(1), 79-90.
- Pérez, E. R., & Medrano, L. A. (2010). Análisis factorial exploratorio: Bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 2(1), 58-66.
- Pérez-Gil, J. A., Moscoso, S. C., & Rodríguez, R. M. (2000). Validez de constructo: El uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*, 12, 442-446.
- Peró, M., Soriano, P. P., Capilla, R., i Olmos, J. G., & Hervás, A. (2015). Questionnaire for the assessment of factors related to university degree choice in Spanish public system: A psychometric study. *Computers in Human Behavior*, 47, 128-138.
- Ramsay, J. O. (1973). Effects of number of categories in rating scales on precision of estimation of scale values. *Psychometrika*, 38, 513-532.
- Revelle, W. (2017). *An introduction to the psych package: Part II Scale construction and psychometrics*. Retrieved from <https://cran.r-project.org/web/packages/psych/vignettes/overview.pdf>
- Revelle, W., & Revelle, M. W. (2017). Package 'psych'.
- Rosseel, Y. (2011). *Lavaan: An R package for structural equation modeling and more Version 0.4-9 (BETA)*. Retrieved [http://byrneslab.net/classes/lavaan_materials/lavaan Introduction4-9.pdf](http://byrneslab.net/classes/lavaan_materials/lavaan%20Introduction4-9.pdf)
- Ruiz, M. A., Pardo, A., & San Martín, R. (2010). Modelos de ecuaciones estructurales. *Papeles del psicólogo*, 31(1).
- Ruiz R. (2012, junio 30). Adiós a la Selectividad. *La Razón*. Recuperado de http://www.larazon.es/historico/357-adios-a-la-selectividad-JLLA_RAZON_470190
- Soriano Jiménez, P. P. (2016). Un modelo estructural para el análisis de los factores asociados a la elección de estudios universitarios (Tesis doctoral no publicada). Universitat Politècnica de València. doi:10.4995/Thesis/10251/64076
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. New York, NY: Harper & Row.
- Universidad de Alicante. (2011). *Reducción de datos: Análisis de componentes principales y factorial*. Recuperado de <http://web.ua.es/es/lpa/docencia/practicas-analisis-exploratorio-de-datos-con->

[spss/practica-5-analisis-multivariante-con-spss-reduccion-de-datos-analisis-de-componentes-principales-y-factorial.html](https://www.spss.com/practica-5-analisis-multivariante-con-spss-reduccion-de-datos-analisis-de-componentes-principales-y-factorial.html)

Vega, L. E. S., & García, L. F. (2009). Dificultades en el proceso de toma de decisiones académico-profesionales: el reto de repensar la orientación en Bachillerato.

Revista de Educación, 350, 323-350.

Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64, 956-972.

El ACP inicial se realizó mediante la extracción de tantos componentes como variables involucradas en el análisis, veinticinco componentes, el resultado se aprecia en la Tabla 10, 11 y 12.

```

## Principal Components Analysis
## Call: principal(r = dtotalest, nfactors = 25, rotate = "none", covar = TRUE)
## Standardized loadings (pattern matrix) based upon correlation matrix
##
##      PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9  PC10  PC11
## p_12      -0.32
## p_13
## p_14
## p_15      0.36
## p_29  0.39      0.38  -0.50
## p_16      0.53  -0.42
## p_30      0.45
## p_17      0.49  -0.46  0.33
## p_31  0.42      0.33  0.34
## p_18      0.34      -0.32      0.31      0.38
## p_32  0.70
## p_19      0.46
## p_33  0.68
## p_20      0.62  0.35
## p_34  0.80
## p_21      0.64
## p_22      0.48  0.34
## p_36      0.53
## p_23      0.72      -0.32
## p_37  0.69
## p_24      -0.66      0.38
## p_38  0.70      -0.38
## p_25      0.57
## p_39  0.73      -0.36
## p_26      0.46      0.38  -0.38

```

Tabla 10. Matriz ACP con 25 componentes (parte I)

En la Tabla 11 se puede apreciar los componentes que agrupan más ítems con los pesos más altos, mientras que en las Tablas 12 y 13 los ítems presentan valores poco representativos en cada uno de los componentes.

##	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22
## p_12	0.48	0.34									
## p_13	-0.57										
## p_14											
## p_15				-0.33							
## p_29								0.35			
## p_16									0.37		
## p_30	0.36	-0.44									
## p_17										-0.35	
## p_31		-0.43									
## p_18											
## p_18				-0.43							
## p_32											0.48
## p_19					0.55						
## p_33											
## p_20								0.42			
## p_34											
## p_21											
## p_22											
## p_22											
## p_36		0.38		0.36							
## p_23											
## p_37											0.36
## p_24											
## p_38											
## p_25							0.41				
## p_39											
## p_26											-0.32

Tabla 11. Matriz ACP con 25 componentes (parte II)

##	PC23	PC24	PC25	h2	u2	com
## p_12				1	-2.0e-15	8.4
## p_13				1	2.2e-15	4.8
## p_14				1	7.8e-16	4.9
## p_15				1	1.2e-15	6.0
## p_29				1	-2.2e-16	7.1
## p_16				1	1.2e-15	6.8
## p_30				1	-1.1e-15	6.9
## p_17				1	-4.4e-16	7.2
## p_31				1	8.9e-16	8.9
## p_18				1	-6.7e-16	9.5
## p_32				1	-5.6e-16	3.2
## p_19				1	7.8e-16	6.1
## p_33				1	6.7e-16	4.2
## p_20				1	3.3e-16	4.7
## p_34				1	6.7e-16	2.3
## p_21				1	-8.9e-16	4.4
## p_22				1	-8.9e-16	4.9
## p_36				1	-6.7e-16	7.2
## p_23				1	-3.3e-16	3.2
## p_37				1	1.4e-15	3.9
## p_24		0.38		1	-4.4e-16	4.1
## p_38			0.31	1	1.0e-15	3.5
## p_25				1	6.7e-16	6.5
## p_39			-0.34	1	2.2e-16	3.1
## p_26				1	8.9e-16	8.8
##						
##						

Tabla 12. Matriz ACP con 25 componentes (parte III)

En la segunda matriz que da el resultado del ACP se aprecia los autovalores (SS Loadings), la proporción de la varianza (Proportion Var) y la varianza acumulada (Cumulative Var).

##	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
## SS loadings	3.88	2.24	1.78	1.44	1.32	1.23	1.11	1.09	0.99	0.94
## Proportion Var	0.16	0.09	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04
## Cumulative Var	0.16	0.24	0.32	0.37	0.43	0.48	0.52	0.56	0.60	0.64
## Proportion Explained	0.16	0.09	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.04
## Cumulative Proportion	0.16	0.24	0.32	0.37	0.43	0.48	0.52	0.56	0.60	0.64
##	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
## SS loadings	0.90	0.87	0.79	0.78	0.73	0.72	0.66	0.61	0.56	0.55
## Proportion Var	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02
## Cumulative Var	0.68	0.71	0.74	0.77	0.80	0.83	0.86	0.88	0.91	0.93
## Proportion Explained	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.02
## Cumulative Proportion	0.68	0.71	0.74	0.77	0.80	0.83	0.86	0.88	0.91	0.93
##	PC21	PC22	PC23	PC24	PC25					
## SS loadings	0.50	0.40	0.36	0.32	0.22					
## Proportion Var	0.02	0.02	0.01	0.01	0.01					
## Cumulative Var	0.95	0.96	0.98	0.99	1.00					
## Proportion Explained	0.02	0.02	0.01	0.01	0.01					
## Cumulative Proportion	0.95	0.96	0.98	0.99	1.00					

Tabla 13. Segunda matriz de ACP

De este primer análisis aplicando el criterio de selección de componentes más común, el criterio Kaiser (Loadings mayores a 1), a partir de los loadings podríamos decir que son ocho los componentes a seleccionar.

