

UNIVERSIDAD SANTO TOMÁS
VICERRECTORÍA DE UNIVERSIDAD ABIERTA Y A DISTANCIA –VUAD–
FACULTAD DE EDUCACIÓN
DOCTORADO EN EDUCACIÓN

Análisis de pregrados colombianos en Estadística y Ciencia de Datos para
la sociedad contemporánea.

Giovanni Martínez López

Bogotá

5 de diciembre de 2022

Análisis de pregrados colombianos en Estadística y Ciencia de Datos para
la sociedad contemporánea.

Giovanni Martínez López

Director de la tesis: Dr. Wilson Hernando Soto Urrea.

Codirector de la tesis: Dr. Julio Ernesto Rojas Mesa

Tesis para optar al título de

Doctor en Educación

Bogotá

5 de diciembre de 2022

Tesis aprobada por:

Handwritten signature

Director de tesis.

Julio Ernesto Rojas Mesa

Codirector de tesis.

Jurados:

[Signature]

Dr. Elías Saíd Yung

Conia Helena Castellanos G.

Dra. Sonia Helena Castellanos



Dra. Carolina Mejía Corredor

Agradecimientos y dedicatorias

Señor Jesús: Hace seis años abriste una puerta hacia un camino que no sabía cómo recorrer y que nadie, salvo tú, pudo cerrar. No terminaba de comprobar si estaba preparado para empezar, cuando repentinamente tuve que dar mis primeros pasos. Muchas veces pensé que te habías equivocado. En otras, no sabía cómo seguir. Con eso y todo, usaste personas en mi camino sin los cuales hubiese sido imposible andar:

- Mis directores de tesis, los doctores Julio Rojas y Wilson Soto, con paciencia y sabiduría tendientes al infinito.
- Mi esposa Margie, mi gran compañera de viaje, mi dato más relevante y fiel presidente de mi club de fans. ¡Gracias por tu amor y ánimo!
- Mi hijo Rafael, mi dato no estructurado favorito, un tanque de vida extra, un recordatorio que todo esto valía la pena.
- Mis queridas Abuela Pepas y Tía Piolín, quienes me brindaron su casa, me inyectaron ánimo, oraron por mí, y motivaron a otros a hacerlo, en especial a Cenfol Semillas de Esperanza y Amor.
- A Cenfol PEC Soledad y sus continuas oraciones para coronar este proceso.
- A la Universidad Católica de Colombia y su Departamento de Ciencias Básicas, en cabeza del Doctor Fredy Garay, por la confianza depositada en mí a lo largo de este proceso.
- A mi madre, mis hermanos y en especial a Don Lucas Martínez: siempre fuiste un buen tipo mi viejo, como diría mi hijo: “papá, esto es para ti”.

Pero sobre todo gracias a ti, quien tiene palabras eternas que dan sabiduría al sabio y conocimiento al estudioso (Daniel 2.21). Que los títulos no sirven de nada si no son sazonados con amor. Gracias por recordarme que mejores son muchos en una causa justa que uno solo, que tu gracia alcanza para disimular mi infinita ignorancia y que la puerta que abriste la cerraste con el final que visualicé. Amén.

Tabla de contenido

1.	Planteamiento y Formulación del Problema	1
1.1	Introducción.....	1
1.2	Problema de Investigación	4
1.3	Objetivos	10
1.3.1	General.....	10
1.3.2	Específicos.....	10
1.4	Justificación del problema y de la investigación	10
2.	Perspectiva Histórica del uso del Dato	13
2.1	De la Escritura a los Censos	14
2.2	El Dato y la Revolución Científica	17
2.3	Aritmética Política: los Datos Amplían su Séquito.....	20
2.4	Los Datos a la Conquista de más Contextos	25
2.5	Datos y Guerras Mundiales: las Dos Caras de la Moneda.....	30
2.6	big-data: un Recubrimiento Abierto para la Humanidad.....	34
2.7	Big-Data: Mucho más que big-data.....	38
3.	La Sociedad del Dato y la Educación Estadística	44
3.1	¿Y Cómo Llegamos a Esto?	45
3.2	Dataísmo: el Credo de la Sociedad del Dato	49
3.2.1	Las Facetas del Dataísmo	50
3.2.2	¿Dato, Luego Existo?	55

3.3	El Prosumidor de Datos	60
3.3.1	El Consumidor de Datos y Alfabetización Estadística.....	61
3.3.2	El Productor de Datos y el Pensamiento Estadístico	66
3.4	Otros Conceptos Relacionados con SL y ST	71
3.5	Estadística y Matemática: ¿Estadísticamente Independientes?	75
3.5.1	Educación Estadística y Educación Matemática: ¿Estadísticamente Independientes?	78
3.5.2	Enseñanza de Estadística: ¿Otro Curso de Matemáticas?.....	79
3.5.3	El Profesor de Estadística: ¿Preparado para Enseñar?	81
3.5.4	¿Cómo son los Resultados Finales de Aprendizaje en Estadística en Educación Superior?.....	83
3.6	Ciencia de Datos: ¿Otro Espacio para el Análisis de Datos?.....	85
4.	Metodología.....	92
4.1	Enfoque de la Investigación	92
4.2	Método de Investigación	94
4.3	Diseño metodológico.....	96
4.3.1	Recolección de información	96
4.3.2	Codificación	97
4.3.3	Sobre la Primera Unidad de Análisis	98
4.3.4	Sobre las Unidades de Análisis Posteriores	101
4.3.5	Triangulación Intra-método y Triangulación de Datos.....	103

4.4	Sujetos de la Investigación e Información Disponible en la Web	104
5.	Resultados	114
5.1	Manifestaciones de la Sociedad del Dato según PECD.....	114
5.1.1	Programas de Estadística.....	115
5.1.2	Programas en Ciencia de Datos	120
5.1.3	Programas de Ingeniería en Ciencia de Datos	124
5.2	El Profesional del Trabajo con Datos según PECD	126
5.2.1	El Estadístico como Profesional del Dato	127
5.2.2	El Científico de Datos como Profesional del Dato.....	141
5.2.3	¿Qué es un Ingeniero en Ciencia de Datos?	151
5.3	El Análisis de Datos en PECD	154
5.3.1	Según los Programas de Estadística	155
5.3.2	Según los Programas de Ciencias de Datos.....	158
6.	Discusión y Conclusiones.....	160
6.1.	Acerca de las Manifestaciones de la Sociedad del Dato.....	160
6.1.1	Big-Data: ¿Sólo Volúmenes Masivos de Datos?	160
6.1.2	Ética en el Trabajo con Datos: ¿Justicia o Beneficios?	162
6.1.3	Estadística y Ciencia de Datos: ¿Discontinuidad Evitable?	164
6.2.	Acerca del Profesional del Dato	167
6.3.	Acerca del Análisis de Datos.....	170
6.4.	Conclusiones.....	171

Referencias bibliográficas	177
----------------------------------	-----

Tabla de figuras

Figura 1	99
Figura 2	102
Figura 3	107
Figura 4	107
Figura 5	116
Figura 6	122
Figura 7	127
Figura 8	128
Figura 9	129
Figura 10	130
Figura 11	139
Figura 12	146
Figura 13	149
Figura 14	149
Figura 15	154
Figura 16	155
Figura 17	156
Figura 18	158
Figura 19	168

Índice de tablas

Tabla 1	108
Tabla 2	109
Tabla 3	110
Tabla 4	112
Tabla 5	118
Tabla 6	144
Tabla 7	147
Tabla 8	150
Tabla 9	152

Lista de siglas

Abreviatura	Significado
ACOFI	Asociación Colombiana de Facultades de Ingeniería
AD	Análisis de Datos
AMI	Alfabetización Mediática e Informativa
ASA	American Statistical Association
ASQ	American Society of Quality
AEC	Alfabetización Estadística Crítica
CUAC	Corporación Universitaria Autónoma del Cauca
ECCI	Universidad ECCI
EXIM	Examen de Ciencias Básicas
FUL	Fundación Universitaria los Libertadores
GAISE	Guidelines for Assessment and Instruction in Statistics Education
GJ	Escuela Colombiana de Ingeniería Julio Garavito
ICD	Programas de pregrado en Ingeniería de Ciencia de Datos
ICFES	Instituto Colombiano para el Fomento de la Educación Superior
INCCA	Universidad INCCA de Colombia
ISI	International Statistical Institute
OCDE	Organización para la Cooperación y Desarrollo Económico
ONU	Organización de las Naciones Unidas
PCD	Programas de pregrado en Ciencia de Datos
PE	Pensamiento Estadístico
PECD	Programas de Pregrado en Estadística y Ciencia de Datos

Abreviatura Significado

PEP	Proyecto Educativo del Programa
PNUD	Programa de las Naciones Unidas para el Desarrollo
QS	Quantified-Self
RE	Razonamiento Estadístico
SL	Alfabetización Estadística
SNIES	Sistema Nacional de Información de la Educación Superior
SOLO	Structure of Observed Learning Outcome
SP	Solución de Problemas
ST	Pensamiento Estadístico
UNICEF	Fondo de las Naciones Unidas para la Infancia
TE	Trabajo con profesionales de otras disciplinas y saberes
TD	Toma de Decisiones
TF	Teoría Fundamentada
UA	Universidad de Antioquia
UB	Universidad el Bosque
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
USTA	Universidad Santo Tomás
UV	Universidad del Valle

Resumen

Si se pudiera describir el escenario actual de desarrollo de la humanidad, con un alto grado de confianza puede decirse que es la Sociedad del Dato. Ante los hechos que señalan el uso continuo y acelerado de los datos, se ha reconocido desde diversos sectores su potencial para impulsar el bienestar social y hasta cierto punto, el conocimiento que tenemos de nosotros mismos. A pesar de que existan voces que cuestionen su uso ciego, nadie escapa a tener que enfrentarse a un mundo cada vez más moldeado por los datos. Por esto, el análisis de datos se ha vuelto indispensable para asegurar que las personas sepan cómo hacer frente a su uso y a la información que pueda generarse de ellos. Esta investigación pretende mostrar las percepciones que los programas de pregrado en Estadística o Ciencia de Datos colombianos tienen con relación al análisis de datos. A partir de la revisión documental de las páginas web de las universidades oferentes, se aplica Teoría Fundamentada con momentos de triangulación intra-método para identificar cómo la Sociedad del Dato se manifiesta en sus propuestas, caracterizar el profesional que están visualizando hacia futuro e identificar el lugar que le otorgan al análisis de datos.

Palabras clave: Dato, Sociedad del Dato, Teoría Fundamentada, Estadística, Ciencia de los Datos, Análisis de datos.

Abstract

Statistically speaking, with a high rate of confidence, Data Society is the scenario of humanity's development. Due to the facts that reveal that data use is more continuous and accelerated, diverse sectors have recognized their potential to improve social goodness and self-knowledge at a certain point. Nevertheless, voices question its blind use, no one escapes having to face a data-centric world. Because of that, data analysis is indispensable to ensure people know what to do with data and information related to them. This research pretends to show how Colombian degrees in statistics or data science programs perceive data analysis. From a documental review of their pages, the use of Grounded Theory with triangulation intra-method allows them to identify what kind of elements they recognize about Data Society, characterize the kind of professional they are visualizing and identify the place that data analysis takes.

Key words: Data, Data Society, Grounded Theory, Statistics, Data Science, Data Analysis.

1. Planteamiento y Formulación del Problema

“Data, data, data, data, data, data, data

Cómo se bebe de una catarata” (Jorge Drexler, Data Data)

1.1 Introducción.

En el corazón de la humanidad aún reposa la necesidad de contar con mecanismos que le permitan explicar los fenómenos que ocurren en el mundo que la rodea. Para ello, y dependiendo de su momento histórico, se ha provisto de recursos con los que se ha logrado, en mayor o menor medida, adquirir conocimiento de sí misma. No importa si se trata de tablas de arcilla, grandes cantidades de rollos de papel, computadores o cualquier dispositivo electrónico que quepa en la imaginación humana; tampoco si le ha tomado poco o mucho de su tiempo para recrear lo observado; el hombre encontró en los datos el retrato perfecto capaz de ilustrar fielmente lo que pasa a su alrededor.

De actor de reparto a protagonista, de gregario a capo de filas, de ser parte de la humanidad a contenerla, el recorrido de la historia muestra, con un nivel cada vez más alto de confianza, la omnipresencia del dato en la actividad humana, premisa en la que diferentes sectores del aparato social y productivo han sembrado su esperanza a cambio de conseguir mayor riqueza y evolución. En ese sentido, el dato se está convirtiendo en la fuente de poder y conocimiento; entender las implicaciones de este contexto es fundamental en la construcción de propuestas educativas orientadas a la formación de profesionales.

Las implicaciones de su uso no solo alcanzan las esferas productivas de la actividad humana, rozan las inquietudes más relacionadas con la existencia y el propósito del hombre en el mundo. El paso del tiempo ha comprobado que enfrentarse a una realidad basada en datos no es un asunto exclusivo de expertos, el asunto está tocando a cada vez más millones de personas al punto de volverse inevitable la necesidad que la sociedad esté provista de capacidades para el análisis de datos. Por esta razón, se hace preciso revisar en qué medida los diferentes escenarios de formación académica están respondiendo a las demandas de una sociedad cada vez más permeada por los datos.

La ciencia que ha dotado de herramientas al hombre para el tratamiento de los datos ha sido la Estadística, fruto de la confluencia entre la aritmética política y la teoría de probabilidades. Por ello, la cuestión relacionada con saber qué hacer con los datos ha sido más investigada en el ámbito de la Educación Estadística que en cualquier otro, y el foco de sus discusiones se ha concentrado en los procesos de enseñanza y aprendizaje de la Estadística.

No obstante, a pesar de que la Educación Estadística se ha desarrollado como campo de investigación, se han levantado voces cuestionando el alcance de los métodos estadísticos para el tratamiento de datos, generando controversias entre los llamados analistas de datos y los estadísticos. Ante la proliferación de los datos no estructurados, la Ciencia de los Datos se ha levantado como un campo emergente para proveer un marco de formación de competencias en análisis de datos. Más allá de abogar por la Estadística o la ciencia de los datos, la pregunta por saber qué hacer con los datos cobra más vigencia como objeto de estudio en la educación.

Esta investigación pretende revelar cómo es visto el análisis de datos por parte de los programas de pregrado de Estadística o Ciencia de Datos en Colombia, y así justificar la urgencia de que las personas adquieran habilidades para el análisis de datos. Para ello, la tesis propone inicialmente un recorrido histórico, no exhaustivo, que permita visualizar cómo los datos han ganado importancia para entender cómo se ha configurado la Sociedad del dato. Acto seguido, se caracteriza esta nueva manera de ver el mundo, cuáles son las implicaciones que tiene para la existencia humana y cómo se ubica al hombre en este nuevo orden.

Ya que la Sociedad del Dato propone nuevos desafíos para la humanidad, posteriormente se muestran los relacionados con el ámbito educativo. Así las cosas, se dirige la mirada a la Educación Estadística y la ciencia de los datos para revisar cuáles son las recomendaciones que investigadores en estos campos de conocimiento han formulado para que los futuros profesionales tengan habilidades en el análisis de datos.

Finalmente, considerando que esta tesis propone explorar los desafíos que la Sociedad del Dato propone en la educación superior, se propone que el análisis de datos debe ser una competencia transversal a todo plan de estudios de pregrado. Por esta razón, se formularán lineamientos que contribuyan a la adquisición de dicha habilidad, y que orienten la construcción de propuestas pedagógicas que produzcan profesionales competentes en términos del tratamiento de datos, y conscientes de las implicaciones que este tiene para la sociedad.

1.2 Problema de Investigación

Si se asume una mirada data-céntrica de la sociedad entonces es muy importante garantizar que las personas posean capacidades para el análisis de datos. El problema es que los profesionales de la sociedad del dato no cuentan con habilidades para el análisis de datos, lo cual se puede explicar por la poca importancia que estas tienen en los currículos universitarios. A pesar de que la inclusión de la Estadística en el currículo universitario ha contribuido con la causa del análisis de datos, hay serias dificultades en los procesos de enseñanza y aprendizaje de la Estadística.

Así mismo, parece que la formación de estas competencias queda relegada únicamente a los semestres donde se ofertan cursos de Estadística y no se hace de manera constante en los períodos siguientes. Como consecuencia de esto, los estudiantes universitarios no siempre cuentan con los elementos que harían de ellos personas competentes en el marco de la sociedad del dato. En esta sección, se explicará en forma sintética las razones de esta problemática.

La Estadística es la disciplina que ha sido más afín al análisis de datos, por tanto, revisar los procesos de enseñanza y aprendizaje en ella se vuelve obligado para entender el problema. Su inclusión en los currículos escolares y un grupo de pregrados profesionales han hecho de la Educación Estadística un campo de investigación en vía de desarrollo (Ramos, 2019). Uno de los mayores esfuerzos que ha asumido la investigación en Educación Estadística es convencer a la comunidad académica que la Estadística y la matemática son ciencias independientes en términos de sus fundamentos y métodos, no obstante, el

impacto de estos esfuerzos no ha sido el esperado y persiste la creencia que la Estadística es una rama de la matemática.

Parte de la situación anterior radica en que la Educación Estadística parece un campo fragmentado e invisible para la comunidad académica en general. Garfield y Ben-Zvi (2008) reconocieron que esto se debe a que las investigaciones relacionadas con el área aparecen más en publicaciones relacionadas con otros campos de conocimiento --Estadística, educación, psicología, tecnología educativa y la educación matemática-- que en la Educación Estadística. Con relación a la educación matemática, el campo se encuentra en un proceso de maduración como disciplina independiente, gracias a su inclusión en currículos, de manera independiente al currículo matemático.

Las cuestiones más importantes de la Estadística son diferentes a las de la matemática. Por ejemplo, el abordaje del concepto de dato en Estadística -- lugar desde el cual se desarrolla esta investigación-- incluye cuestionar su proceso de producción y valorar el contexto que está representando, lo cual no ocurre en la matemática. Por ejemplo, Moore (1997) hizo referencia a la formulación de ejemplos, y argumentó que en Estadística es necesario mostrar una conexión entre el contexto aplicado y el patrón que da lugar a los datos, lo que en matemática no tiene tanta importancia.

Las dificultades en torno a los procesos de enseñanza y aprendizaje radican en no tomar en cuenta las diferencias entre lo importante en Estadística y en matemática. Sin importar el nivel de escolaridad que se considere o el contexto sociocultural, la enseñanza de conceptos estadísticos ha adoptado un carácter

determinista (Zapata, 2011, 2016; Weiland, 2017). Más concretamente, en el contexto de la educación superior colombiana, Rocha (2012) mencionó que la prevalencia del pensamiento determinístico en quienes enseñan probabilidad y Estadística es un obstáculo para mostrar ideas centrales de la Estadística como la aleatoriedad y el manejo de la incertidumbre.

Si se revisan investigaciones alrededor de quienes enseñan Estadística, el panorama no resulta favorable para la formación de competencias en análisis de datos. Con relación al profesor de Estadística, se encontró que no siempre cuenta con la preparación para asumir de manera adecuada cursos de Estadística, ni el conocimiento disciplinar suficiente para formar a los estudiantes en habilidades para el análisis de datos.

A pesar de que los profesores de Estadística, tanto en escuelas como universidades, provienen de una amplia variedad de perfiles profesionales (Rocha 2012; Estrella, 2017; Ramos, 2019; Sierra-Bonilla y Rodríguez, 2019), esto dificulta los procesos de aprendizaje y enseñanza en el aula debido a la falta de preparación que suelen manifestar en términos didácticos, principalmente por un predominio de pensamiento determinista que hace difícil alcanzar una comprensión adecuada de lo aleatorio.

Así mismo, se ha encontrado que quienes enseñan Estadística la usan de forma limitada en su cotidianidad y comparten nociones erradas con los estudiantes sobre los conceptos más importantes (Watson, 2001; Makar y Confrey, 2005; Stohl, 2005). Adicionalmente, Estrella (2017) advirtió sobre la falta de familiaridad de los profesores con el carácter no determinista del análisis

de datos, lo que constituye un obstáculo para la formación de esas habilidades en los estudiantes en términos de la sociedad del dato; después de todo, si los profesores no cuentan con competencias para el análisis de datos es imposible que sus dirigidos las tengan.

Con relación a los estudiantes universitarios se perciben problemas relacionados con sus resultados de aprendizaje, que en parte tiene su explicación en sus actitudes hacia la Estadística. Aunque, gracias a la inclusión de contenidos estadísticos en el currículo escolar es altamente probable que hayan tenido experiencia previa con conceptos estadísticos y probabilísticos en su formación escolar, persisten actitudes negativas hacia la Estadística asociadas con la falta de habilidades matemáticas, el uso que puedan brindarle en su cotidianidad y su interés por estudiarla. Lo anterior resulta paradójico si se considera que la comunidad académica implicada en los procesos de enseñanza y aprendizaje es parte de la Sociedad del Dato.

Justificando lo anterior, Pérez et. al. (2015) hicieron un estudio sobre las actitudes de estudiantes universitarios en una universidad privada, en el que se reportó que no se sienten seguros de sus habilidades en Estadística, no les gusta, ni están interesados en estudiarla. García y Barrero (2018), en un estudio similar al anteriormente mencionado, ratifican las anteriores conclusiones y añaden que no tienen claro cómo usar la Estadística en contextos cotidianos y profesionales. Ramos (2019) observó que, a pesar del reconocimiento de su utilidad, las actitudes desfavorables hacia la Estadística interfieren con los procesos de aprendizaje y pueden asociarse con sentimientos negativos hacia las actividades con números.

Como consecuencia de estas dificultades anteriormente, los estudiantes que finalizan sus procesos de formación profesional evidencian falencias en competencias relacionadas con el análisis de datos como el razonamiento cuantitativo y el razonamiento científico. En cada uno de los módulos mencionados hubo un porcentaje muy reducido de personas que presentaron la prueba, que obtuvo puntajes dentro de los niveles más altos de desempeño, algo que se acentúa más en las universidades privadas (ICFES, 2020a, 2020b).

Ahora bien, el lugar que ocupa la formación Estadística para pregrados en general se concentra en los cursos formales de Estadística descriptiva, probabilidad o Estadística inferencial, pero no hay evidencias de si estos elementos se trabajan en semestres posteriores a los de estas asignaturas. Inclusive, hay todavía carreras en las que ni siquiera se ofrecen cursos de Estadística, con lo que se acorta la probabilidad de que sus egresados tengan habilidades en el análisis de datos. Lo anterior merece cuidado porque el análisis de datos es una competencia necesaria en un rango cada vez más amplio de profesiones, incluyendo aquellas donde no se pensaría en su necesidad. (Oppenheimer, 2018). Así las cosas, los currículos universitarios se están quedando cortos con su aporte a la causa del análisis de datos.

La situación puede considerarse más crítica si se toma en cuenta la emergencia de la ciencia de los datos, algo de lo que se empezó a hablar hacia la mitad del siglo XX y que es otro lugar donde se reclama la necesidad del análisis de datos. El impacto de las controversias entre los científicos de datos y estadísticos no solo han afectado la labor de estos últimos, también han generado cuestionamientos sobre los planes de estudio en pregrados de Estadística,

incluso, se ha llegado a decir que las técnicas Estadísticas ya no resultan tan eficientes para tratar los datos que la humanidad genera actualmente (Gould, 2010; Hicks e Irizarry, 2018).

A manera de resumen, lo explicado en los párrafos anteriores presenta retos para las universidades en términos de las demandas de la emergente sociedad del dato. El panorama actual de la enseñanza de la Estadística en educación superior muestra que se requieren transformaciones de fondo en los procesos de enseñanza y aprendizaje. Sin embargo, la expansión de la Ciencia de Datos abre el debate de si la Estadística es suficiente para lograr habilidades en el análisis de datos, lo que las obliga a estar más abiertas a la posibilidad de dar un lugar más amplio para ellas.

Habrá que revisar cómo se conceptualiza el análisis de datos desde las universidades que ofertan pregrados en Estadística o Ciencia de Datos – de aquí en adelante PECD–, no solo porque son los dos campos de conocimiento más compatibles con el análisis de datos, sino porque esta descripción puede ser el punto de partida para que desde la educación superior se puedan identificar estrategias que ayuden a formar a los profesionales inmersos en la Sociedad del Dato, con las habilidades suficientes en análisis de datos.

Así las cosas, ¿Cuáles son las interpretaciones que los PECD hacen sobre el análisis de datos?

1.3 Objetivos

A continuación, se mencionarán los objetivos de esta investigación, el general y los específicos.

1.3.1 General

Explicar, bajo los parámetros de la TF, las interpretaciones que los PECD hacen sobre el análisis de datos, en el marco de la emergente Sociedad del Dato.

1.3.2 Específicos.

- Describir los elementos que se manifiestan de la Sociedad del Dato a partir del análisis de información sobre los PECD colombianos.
- Caracterizar el futuro profesional en análisis de datos que proyectan las universidades colombianas que ofertan PECD.
- Identificar el lugar que otorgan los PECD colombianos al análisis de datos en la formación de profesionales expertos en el manejo de datos.

1.4 Justificación del problema y de la investigación

Para entender las razones por las que esta investigación se lleva a cabo, vale la pena hacer una exposición de la importancia que el dato ha adquirido en el desarrollo de la humanidad y cómo este brinda una perspectiva diferente para observar la realidad. Esta presentación es necesaria para hacer notar que el dato no solo ha sido un elemento de interés para grupos reducidos de personas o que es de uso exclusivo en un selecto grupo de contextos; que más allá de la postura

frente a ellos o la afinidad hacia su manejo, toda persona accede a ellos de manera libre, y que su actitud hacia ellos es capaz de determinar su comportamiento ante el mundo.

Así mismo, al destacar los factores sociales que le han dado forma, se puede comprender el mundo desde una visión data-céntrica y el rol que el hombre desempeña ante la misma; pues esta concepción del mundo levanta cuestionamientos sobre la relación del hombre con el flujo de datos al que alimenta y se ve expuesto, pero que por momentos parece superarlo. Los datos que genera la humanidad pueden llegar a revelar sobre ella mucho más de lo que se cree y este conocimiento está disponible para cualquiera, más allá que esto pueda representar peligros para la intimidad misma de las personas. Al poner de manifiesto los factores sociales, se busca generar reflexión acerca de cómo enfrentarse a un mundo mediado por los datos sin ser simplemente un aportante en el agobiante flujo de datos de la actualidad.

Por otro lado, los datos se presentan como una forma de lenguaje que es capaz de vincular diferentes campos de conocimiento hasta el punto de romper las barreras existentes entre ellos. Siendo las universidades los lugares por excelencia de generación de nuevo conocimiento, es relevante conocer las consecuencias derivadas de la Sociedad del Dato en la educación superior porque su conocimiento puede contribuir a configurar un perfil profesional a la medida de las demandas de la Sociedad del Dato; no solo pensando en que el futuro egresado sea capaz de usar los datos como herramienta para generar conocimiento, sino que reflexione acerca de las implicaciones éticas y los problemas derivados de su uso.

Cuando se menciona que la investigación busca formular lineamientos para hacer del análisis de datos un componente transversal a la formación de profesionales, es porque se pretende que sean una herramienta con la que se elaboren propuestas curriculares más tendientes al desarrollo de competencias relacionadas con el análisis de datos, pero también de la sensibilidad necesaria para estar más persuadidos de sus peligros. Si bien, se puede pensar que esto es responsabilidad de quienes orientan cursos de Estadística o afines, esta es una labor compartida entre decanos, directores de departamentos, e incluso profesores lejanos a cursos de carácter cuantitativo, dado que toda persona debe tener competencias básicas para el análisis de datos.

2. Perspectiva Histórica del uso del Dato

“Data, data, give me data. I can’t make bricks without straw (Datos, datos, d nme datos. No s  hacer ladrillos sin paja)”.

— Sherlock Holmes

En este cap tulo se desarrollar n aspectos relacionados con el uso del dato a partir de la invenci n de la escritura hasta nuestros d as. Aunque la revisi n no es exhaustiva, se busc  recopilar suficientes hechos y organizarlos cronol gicamente para que se aprecie c mo el dato ha cobrado una relevancia cada vez mayor en el desarrollo de la humanidad.

Principalmente, la actividad humana con los datos ha girado alrededor de dos asuntos. El primero de ellos, encontrar formas en las que estos puedan ser almacenados o procesados para sacar provecho de ellos. El segundo, convertir estos datos en informaci n que contribuya a mejorar el conocimiento de los fen menos representados a trav s de ellos.

A partir de esta revisi n hist rica que se propone, el lector podr  constatar que la humanidad ha podido encontrar, independientemente de la  poca, maneras de procesar datos, sin querer decir con ello que el asunto no haya sido complicado. Sin embargo, el hecho m s importante es que con el desarrollo de la historia, es cada vez m s dif cil encontrar fen menos que no sean representables a trav s de datos.

2.1 De la Escritura a los Censos

Procesar información derivada de las actividades humanas es una necesidad que ha convivido con el hombre. Si bien, las primeras comunidades se caracterizaban por ser pequeñas en tamaño y el cerebro humano podía recoger datos fácilmente, con el inicio de la revolución agrícola, hacia el 9000 a. C. empiezan a surgir civilizaciones cada vez más grandes, lo que llevó a un aumento considerable de datos por procesar. Ante el aumento de información por procesar, el ser humano no estaba preparado para hacerlo por sí mismo, más aún cuando se trataba de datos numéricos.

Harari (2014) explicó que el cerebro humano no puede almacenar vastas cantidades de datos por tres razones: porque su capacidad para hacerlo es limitada por más que pueda llegar a tener habilidades extraordinarias de memoria; porque siempre habrá pérdida parcial o total cuando se intenta replicar memorias de un cerebro a otro; y porque no está adaptado a ciertos formatos, entre ellos, el numérico. Por lo anterior, el hombre requería de un sistema de almacenamiento externo a su cerebro con el que pudiese captar cantidades más grandes de datos.

Adicionalmente, Harari (2014) mencionó que la primera civilización que entendió la importancia de superar tal limitación fue la sumeria, para lo cual inventaron la escritura, una manera de controlar sus actividades comerciales. “Firmado, Kushim” era lo que decía el final de una tablilla de arcilla encontrada hacia 3400 a. C. que probablemente contenía el registro de la cantidad de cebada recibida por un individuo a lo largo de 37 meses. El primer ejercicio de escritura

conocido por la humanidad es un ejemplo de cómo el hombre empezó a registrar datos de sus actividades.

La historia de las civilizaciones muestra diferentes casos de civilizaciones aprendieron a generar sistemas de procesamiento de datos análogos y formas de escritura más completas que la inventada en Sumer. Harari (2014) encontró ejemplos como los quipus de las culturas precolombinas para registrar actividades comerciales, y mencionó sistemas de escritura completos como los de Mesopotamia, China y Egipto, con los que se registraban datos numéricos y se daba lugar a creaciones literarias. Las diferentes formas de escritura ayudaron a solventar las limitaciones humanas para procesar principalmente datos numéricos.

La confianza delegada en los sistemas de escritura hizo posible retratar en forma más adecuada la actividad humana, y entre otras cosas, contar con datos sobre poblaciones más grandes. Es así como, algunas civilizaciones empiezan a realizar censos, esto es, recuentos regulares de las poblaciones a partir de la observación de variables demográficas y actividades económicas. En Quintela del Río (2019) se menciona que los egipcios realizaron un censo hacia el 3050 a. C. en el marco de la Dinastía I, posiblemente para delegar responsabilidades en la construcción de pirámides; también, que en China se hizo un primer recuento de la población bajo el mando del emperador Yao con fines comerciales, agrícolas e industriales.

En la Biblia hay varios ejemplos de censos. Por ejemplo, hacia 1500 a. C. se realizó un censo de las tribus de Israel con propósitos militares para establecer cuántos hombres eran aptos para la guerra (Números 1:3, Nueva Versión

Internacional), prácticas que el rey David realizó con idéntico propósito más adelante (1 Crónicas 21: 1-4, Nueva Versión Internacional). El rey Salomón también realizó un censo de extranjeros para asignar responsabilidades en la construcción del templo de Jerusalén (2 Crónicas 2.17, Nueva Versión Internacional) y el emperador Cesar Augusto ordenó un censo en todo el Imperio Romano en tiempos de Cirenio (Lucas 2.1, Nueva Versión Internacional) con el fin de consolidar una base para recaudar impuestos.

Los romanos fueron quienes hicieron de los censos una práctica periódica, llevándolos a cabo cada cinco años (5 a. C.- 73 d. C). Para ello, crearon la figura del censor, quien tenía la responsabilidad de hacerlos, pero también de vigilar moralmente al imperio y recaudar impuestos (Censo, 2019). Los datos recogidos de los censos republicanos sobre los ciudadanos eran importantes para determinar ciertos derechos y deberes establecidos por las legislaciones imperiales (Mateo, 2016). Con el tiempo se dejaron de nombrar censores y los recuentos de las poblaciones pasaron a ser responsabilidad de los emperadores de turno.

Durante la Edad Media, continuaron los ejercicios de recolección de datos con los empadronamientos (censos municipales) y catastros (descripciones de los inmuebles) en otras partes del mundo. En Censo (2019) se mencionaron casos como la descripción numérica de la Península Ibérica hacia 721; el Libro Domesday en 1086, orientado a recopilar datos de lo que poseía cada dueño de una porción de tierra en aquella población para determinar cuánto podía tributar; diversos ejercicios de empadronamiento descritos en las actas de las cortes de

Alcalá, hacia 1348; y lo hecho por Felipe II, que consta de 636 relaciones de los 13000 pueblos de la Península Ibérica.

Como se puede apreciar, contar con datos de calidad y saber qué hacer con ellos, se convirtió en un asunto de gobierno porque era importante para el manejo de recursos y tomar decisiones en lo económico, lo militar y lo social. Saber lo que se debía hacer con relación a una población dependía del almacenamiento y procesamiento de los datos. Los censos ofrecen un ejemplo práctico de cómo los datos empiezan a ser una herramienta para extraer conocimiento de cómo podrían funcionar las sociedades de una mejor manera, pero esto es apenas una faceta dentro de una serie de inquietudes importantes para su existencia.

2.2 El Dato y la Revolución Científica

En tiempos como la Edad Antigua o la Edad Media, para resolver las cuestiones más interesantes de la humanidad bastaba con la ayuda de un superior. Así como los gobernantes de turno tenían a su disposición conjuntos de datos sobre sus compatriotas y con ello tomar decisiones sobre ellos, las autoridades religiosas tenían respuestas a las preguntas más relevantes de la época, aún si contaran con datos o no para elaborar sus conclusiones. Todo lo que era importante, digno de investigar o conocer dependía del concepto de sabios, dioses o autoridades reinantes; lo que ellos no supieran se consideraba irrelevante para la sociedad.

En realidad, la concepción religiosa de esos tiempos determinó los intereses investigativos de los pensadores de la época; con lo que el acceso al

conocimiento surgía de la explicación que se pudiera brindar de la naturaleza a partir de las tradiciones religiosas. Por ejemplo, Guerra, Macedo y Carvalho (2006) mencionaron que las preocupaciones más relevantes de dos de los pensadores medievales más importantes -Agustín de Hipona y Tomás de Aquino- se ubicaban en cuestiones éticas y de la conducta, así como en la adaptación de los fenómenos naturales a los escritos bíblicos. En general, la importancia de la religión en los asuntos vitales de la sociedad permeó el avance científico de la época, limitando la producción de conocimiento a esferas sociales con altos privilegios económicos.

Harari (2014) explicó que en tales épocas el hombre no creía en su capacidad para adquirir nuevos poderes económicos, militares o médicos, y que los esfuerzos de los gobiernos a nivel educativo se concentraban en mantener las capacidades existentes, no tanto en la adquisición de nuevas habilidades. También, que, aunque había personas que pensaban que existían cosas importantes que la tradición ignoraba, éstas eran marginadas o en el peor de los casos, perseguidas.

Sin embargo, el hombre empieza a descubrir que no todas sus preguntas importantes tenían respuestas. Parafraseando a Bertomeu (2020) en una cita que hace de William Whewell (1840), transitó de confiar en tener todo resuelto en su mente (o en una mente superior a él) a depender de sus observaciones externas sobre las cosas, y de venerar la sabiduría del pasado a esperar pacientemente en el cambio y su expectativa de progreso. Este cambio en la manera de pensar genera una serie de transformaciones en la sociedad europea, que colocaron en

entredicho las concepciones medievales de la naturaleza y determinaron nuevos horizontes para cualquier campo de la actividad humana: la revolución científica.

El producto emblemático de la revolución científica es la ciencia moderna, una forma alternativa de acceder al conocimiento basada en que no basta con revisar los saberes del pasado, y, por lo tanto, deben recolectarse datos empíricos para formular teorías sobre las cosas que se desean conocer. Según Harari (2014) hay tres principios que la gobiernan: (a) la disposición a admitir ignorancia, es decir, que no se conocen todas las respuestas a las preguntas más relevantes ni que ningún conocimiento es infalible; (b) la centralidad de la observación y la matemática, siendo esta última fundamental para formular, a partir de los datos observados, teorías sobre las cosas; y (c) la adquisición para crear nuevos poderes, esto es, desarrollar nuevas tecnologías.

Con la llegada de la revolución científica se da un impulso definitivo a la generación de conocimiento. Si bien, la invención de la imprenta moderna gracias a Johannes Gutenberg (1440) ayudó a la divulgación de ideas alternativas a las medievales, la ciencia moderna proporciona nuevos métodos en los que los datos empíricos y la matemática son protagonistas. De hecho, la matemática se convierte en el lenguaje de las ciencias naturales en lugar de la lógica escolástica imperante en el medioevo, siendo esta la que permite enlazar las observaciones recolectadas de la experiencia, y brindando nuevas posibilidades a campos de conocimiento como la filosofía y la astronomía (Guerra, Macedo y Carvalho, 2006; Harari, 2014).

Todo apunta a que la revolución científica comenzó con un acontecimiento ligado a la astronomía, en 1543 salió a la luz la obra de las revoluciones de las órbitas celestes de Nicolás Copérnico, quien hizo parte de una primera generación de científicos modernos, entre los que se destacan a Johannes Kepler, Isaac Newton y Galileo Galilei, cuyos trabajos trastornaron las ciencias naturales y resaltaron las bondades del razonamiento inductivo (Alfaro, Meseguer, Mondéjar y Vargas, 2011). Sus métodos de investigación son lo que en el sentido moderno se denominó el método científico, una secuencia de etapas que comienza con la observación de un fenómeno de interés, la formulación de preguntas, la elaboración de hipótesis, la experimentación, la demostración de hipótesis, y la formulación de tesis.

Así las cosas, el hombre, al reconocer que no tenía la última palabra sobre todo lo importante para la humanidad, no solo se vale del conocimiento antiguo sino de los datos observados a partir de su propia experiencia para elaborar inferencias acerca de los fenómenos que estudia. Los datos se convirtieron en registros que permiten representar aquello que se puede observar desde la experiencia misma, y la matemática, en el lenguaje que permitía representar lo escondido tras ellos para lograr teorías.

2.3 Aritmética Política: los Datos Amplían su Séquito

El razonamiento inductivo, base del método científico, tuvo un éxito que se tradujo en cambios a nivel económico y tecnológico, lo que llevaría a pensar que sus métodos podrían ser replicables en las ciencias sociales. Continuando con el legado dejado por la primera generación de científicos contemporáneos a

Copérnico, se levantó una posterior cuyo propósito era aplicar técnicas cuantitativas para el tratamiento de cuestiones relacionadas con el hombre y su actividad social.

La revolución científica supuso transformaciones económicas dignas de considerar. El fracaso del feudalismo medieval impuso una nueva forma de concebir la economía, dando paso al capitalismo y a la formación de estados nacionales. Adicionalmente, la visión familiar y aristotélica de la economía medieval quedó reemplazada por una economía mercantil.

La economía, adquiere una dimensión pública, y, por tanto, política, con lo que surgieron nuevas inquietudes a los pensadores económicos; conceptos como precio, mercancía y mercado se convirtieron en objetos de estudio (Guerra, Macedo y Carvalho; 2006). Así mismo, Alfaro, Mesequer, Mondéjar y Vargas (2011) mencionaron que la actividad recaudatoria del estado, “la actividad recaudatoria del estado, las transacciones del comercio exterior, la aparición de la balanza comercial o las necesidades de conocimiento de la población ocasionaron la necesidad de recolectar y estimar datos para fines estatales” (p. 3).

Según Quintela del Río (2019), son los ingleses los primeros en notar que el análisis de datos a lo largo del tiempo permitía detectar patrones con los que podrían describir el comportamiento de la sociedad. Así las cosas, la segunda mitad del siglo XVII vio el surgimiento de esa generación posterior a la de Copérnico, entre los que se destacaban John Graunt, Samuel Hartlib, John Evelyn, John Dury, Robert Boyle y William Petty quienes llevaron el uso de los datos a la exploración de fenómenos no cuantificados. Por ejemplo, John Graunt

sentó las bases de la demografía gracias a *Natural and Political Observations on the Bills Mortality* en 1662. Alfaro, Meseguer, Mondéjar y Vargas (2011)

mencionaron que Graunt notó la presencia de la ley de los grandes números en los fenómenos sociales y usó la regla de tres como herramienta analítica; tal y como lo mostraron a continuación:

“[...] cerca de un tercio de los niños nacidos vivos, muere en los primeros cinco años y alrededor del 36% por debajo de los 6. Sea la peste grande o pequeña, Londres estará completamente repoblada al cabo de 2 años. Hay en torno a 6.5 millones de personas en Inglaterra. En Londres hay 14 varones por cada 13 mujeres. En todo matrimonio, unos con otros, se tiene 4 niños”. (p. 7).

Para Piovani (2007) lo más interesante de la obra de Graunt no reside en la organización o recolección de los datos sino en las conclusiones recogidas a partir de ellos, y la justificación de la necesidad de una disciplina: la aritmética política. Diversos investigadores coinciden que este fue el primer antecedente de desarrollo de la Estadística como disciplina (Yañez, 2000; Piovani, 2007; Quintela del Río, 2019); de hecho, Ferreiro y Fernández (1988) mencionaron que el adjetivo *statisticus* empezó a usarse con frecuencia en el siglo XVII para designar al encargado de los asuntos estatales.

William Petty, en su obra *Political Arithmetick* (1691), introdujo formalmente la noción de aritmética política. En este trabajo, Petty logró abordar los asuntos económicos del Reino Unido mediante la obtención de datos sobre características observables en la población y su posterior análisis, aplicando el método científico.

Alfaro, Meseguer, Mondéjar y Vargas (2011) citaron una parte de su obra para ilustrar el enfoque que propuso:

“en lugar de usar solo comparativos y superlativos y argumentos intelectuales, yo me he propuesto expresarme en términos de números, pesos y medidas, usar solo argumentos que tengan sentido y considerar únicamente las causas que estén visiblemente basadas en la naturaleza, abandonando la consideración de todo aquellos que dependen de la naturaleza mutable de las mentes, las opiniones, los apetitos y las pasiones de los hombres particulares” (p. 3)

Según Guerra, Macedo y Carvalho (2006) la aritmética política es “la extensión de una nueva ciencia social, es decir, la afirmación de métodos cuantitativos en el análisis de los fenómenos sociales, en estrecha sintonía con la revolución científica de las ciencias naturales.” (p. 12). La aritmética política fue el resultado de la recolección, procesamiento y análisis de datos para gobernar bien, y, por tanto, mejorar la sociedad. Tal expresión se convirtió más adelante en el nombre *Estadística Social*, esto es, “una ciencia de la sociedad cuyas conclusiones dependían de números y medidas.” (Piovani, 2007, p. 28).

La aritmética política británica como Estadística social y los trabajos de Petty tuvieron eco en la resolución de problemas vinculados a mejorar la sociedad. Su florecimiento es contemporáneo con la Primera Revolución Industrial, la cual es resultado de transformaciones tecnológicas y económicas que empezaron en Inglaterra hacia la segunda mitad del siglo XVIII. En ese tiempo se gestó el encuentro entre la *political arithmetic* británica y la *universitätsstatistik*, vertiente

alemana de la aritmética política cuyo desarrollo fue contemporáneo a la británica, y que consideraba el tratamiento de los aspectos más relevantes del estado, para lo que ellos no veían la necesidad de contar con datos numéricos (Piovani, 2007).

El encuentro no estuvo exento de controversias vinculadas con los fines de la Estadística como disciplina pues estuvo en juego su finalidad como disciplina, de todos modos, fueron los británicos quienes impusieron sus argumentos a los alemanes en favor de la cuantificación para tratar asuntos del estado. En efecto, la tendencia al uso de datos numéricos en la investigación social cobró fuerza con la construcción de matrices de datos para el análisis comparado de los estados alemanes. Al respecto, Marradi, Archenti y Piovani (2007) explicaron que esta herramienta fue desarrollada por Godofredo Achenwall “poniendo reinos, ducados y ciudades libres en las filas y sus estados en varias propiedades en las columnas de una rudimentaria matriz.” (p. 21), y colocando en el contenido de los cruces expresiones verbales, que posteriormente fueron reemplazadas por registros numéricos.

Así mismo, en el Reino Unido empezaron a gestarse las primeras asociaciones dedicadas exclusivamente a la Estadística, en particular la *Royal Statistical Society* que contaba en sus filas con economistas, políticos, miembros de la cámara de los Lords, representantes del gobierno y médicos, y cuyo interés estaba en entender las transformaciones sociales de la época, sentando bases científicas para las políticas sociales (Piovani, 2007).

Siguiendo a Piovani (2007), los *statists* pensaban que los procesos de urbanización propios de la Primera Revolución Industrial explicaban el carácter y

costumbres de las poblaciones; desde su labor recolectando y organizando datos, estudiaron un tipo de Estadísticas denominadas como “morales” por tratarse de asuntos como la educación y la criminalidad, para impulsar reformas sanitarias y educativas que mejorarían la calidad de vida de los obreros.

Aunque hasta ese momento el tratamiento de los datos no contaba con técnicas matemáticas sofisticadas, se podía visualizar un público cada vez más amplio al que le interesaba contar y procesar datos para fines sociales. Gracias a la confluencia entre las vertientes británica y alemana, los datos poblacionales ampliaron su séquito, dando forma a disciplinas como la Estadística y la demografía. La aritmética política es entonces un testimonio de cómo contar con datos y analizarlos de acuerdo con el método científico se trasladó a cuestiones sociales. Así mismo, es un ejemplo del interés de más personas en el mundo por recoger datos para propósitos cada vez más variados. Obtener datos sobre la sociedad se convirtió en la piedra angular para hacer que las sociedades funcionen mejor.

2.4 Los Datos a la Conquista de más Contextos

Con el paso del tiempo, los procesos de industrialización iniciados en Inglaterra se extendieron al resto de Europa, Japón y Estados Unidos, configurando un escenario globalizante para la economía y dando paso a la Segunda Revolución Industrial, fenómeno que marcó los tiempos de la segunda mitad del siglo XIX y la Primera Guerra Mundial. La tendencia a la cuantificación iniciada por el desarrollo de la aritmética política se consolidó gracias a la incursión de la teoría de probabilidades, brindando a la Estadística el estatus de

ciencia. Mientras el mundo seguía transformando la manera de percibir las actividades cotidianas, el rango de disciplinas en donde era importante contar con datos y analizarlos sería más amplio.

El desarrollo tecnológico característico de la Segunda Revolución Industrial se vio favorecido por el descubrimiento de nuevos metales (acero, zinc, aluminio, níquel, manganeso, cromo, cobre) que impulsaron la industria eléctrica, lo cual se explica en gran parte por los avances en ciencia. La globalización afectó positivamente la divulgación del conocimiento científico, haciendo que su difusión sea universal y animando a las universidades a la formación de personas en ciencias. La ciencia facilitó el surgimiento de nuevas industrias, inventos que mejoraron la calidad de vida de las personas, y empezó a ser de utilidad en la solución de problemas no cuantificables hasta ese momento.

Ejemplo de lo anterior se encuentran en las Estadísticas médicas. Ruiz (2006) menciona casos que datan del siglo XVII, del uso de datos para buscar relaciones entre variables demográficas y de tipo sanitario, pero afirmó que es hacia 1830 cuando las técnicas cuantitativas tomaron partido gracias a los trabajos del clínico francés Pierre Charle Alexandre Louis, considerado como uno de los médicos que más abogó por la cuantificación en la medicina y quien estuvo interesado en estudiar con detalle la tuberculosis, una de las enfermedades más peligrosas en ese entonces.

Fresquet (2007) brindó más detalles de lo realizado por Louis; su estudio llamado *Recherches anatomopathologiques sur la phthisie Précédées du rapport fait à l'Académie royale de médecine par MM Bourdois, Royer-Collard et Chomel*

se basó en 123 observaciones sobre las lesiones y síntomas desencadenados por la tuberculosis pulmonar, para formular la Ley de Louis, según la cual “la tuberculosis pulmonar comienza generalmente por el vértice y en el pulmón izquierdo y, siempre que hay tuberculosis en otras partes del cuerpo, se acompaña también de lesiones en los pulmones” (p. 2). Lo anterior contradecía al supuesto que la tuberculosis se transmitía en forma hereditaria.

Una referencia adicional sobre Louis se encuentra en Morabia (2006) con un estudio publicado en 1836 titulado *Researches On The Effects Of Bloodletting In Some Inflammatory Disease*, acerca de los efectos de la sangría en el tratamiento de la neumonía. Conformó dos grupos de 41 y 36 personas con edades similares; en el primero de ellos, se hizo flebotomía antes de los 5 días posteriores a la enfermedad, y en el otro entre el quinto y noveno día. Observó que en el primer grupo el porcentaje de decesos por la enfermedad fue de 44%, mientras que en el segundo fue 25%, algo que contradecía la evidencia según la cual un tratamiento más pronto reducía las probabilidades de fallecer por neumonía.

Otra área que vio la tendencia a la recolección y el análisis de datos sería la industria, pues las personas con formación científica fueron requeridas para tratar diversidad de asuntos relevantes en esta área. Por ejemplo, Barreto-Villanueva (2012) señaló que el uso de datos y técnicas cuantitativas fue indispensable en las fábricas con el fin de realizar experimentos de bajo costo, para la obtención de nuevos y mejores productos. Bertomeu (2020) complementa lo anterior, mencionando que su uso se extendió a los asuntos del gobierno relacionados con la salud, la guerra, la alimentación y la criminalística; pero que también, las

personas con formación científica empezaron a trabajar en la industria química, agroalimentaria y farmacéutica, y en laboratorios industriales con el fin de racionalizar la producción.

El uso de los datos estaba limitado a su recolección y organización sin llegar a la modelización matemática para la identificación de tendencias presentes en ellos, lo cual se hizo posible gracias a la incursión de la teoría de probabilidades en la Estadística. En efecto, la *political arithmetic* británica y la *universitätsstatistik* alemana cruzaron caminos con la *théorie mathématique des probabilités* francesa, cuyos inicios se remontan a los trabajos de Fermat y Pascal sobre el tratamiento matemático de los juegos aleatorios, y se desarrolló en forma paralela al pensamiento británico y alemán sobre la Estadística.

Según Piovani (2007), hasta antes que la vertiente francesa influyera en el pensamiento estadístico de la época, el tratamiento matemático de los datos poblacionales no gozaba de la fortaleza matemática por estar centrada en la observación de lo típico o normal. La teoría de probabilidades brinda un nuevo paradigma al estudio de datos porque considera a la variabilidad presente en ellos como un fenómeno que debe tratarse.

La difusión de los trabajos de Pierre-Simon Laplace sobre la aplicación de la teoría de los errores en astronomía y el método de los mínimos cuadrados hacia 1827, brindó a la teoría de probabilidades un estatus digno de considerar; desarrollo que también estuvo cimentado en los trabajos de Simeón Dennis Poisson para entender las regularidades matemáticas en contextos cada vez más

amplios, y los de Adolphe Quetelet sobre la ley de los errores (Quintela del Río, 2019).

Quetelet extendió los métodos de la teoría de probabilidades usados en astronomía, a la recolección y clasificación de datos en las cuestiones humanas. Bowler y Morus (2006) mencionan que Quetelet reunió datos poblacionales para revelar la causa principal de los decesos en Francia, descubrió que existían patrones regulares en la manera como se producían actividades ligadas al crimen y el suicidio, y que cualquier característica de la población puede explicarse en términos de la variación respecto a un promedio, concepto clave en la teoría de errores.

El tratamiento que hizo Quetelet de estos datos llamó la atención de Francis Galton quien usó la teoría de probabilidades para abordar asuntos relacionados con la herencia de los caracteres humanos. Motivado por su primo Charles Darwin y la publicación del *Origen de las Especies* hacia 1859, analizó los datos provenientes de la observación de las características de las razas, interesándose por estudiar sus variaciones. Barahona (2005) mencionó que Galton observó atributos como las aptitudes musicales, las preferencias por cierta carrera académica en lugar de otra, y variables antropométricas; y analizando los datos obtenidos encontró que podían deducirse curvas Estadísticas que permitían comparar el comportamiento de las variables observadas a través de diferentes grupos sociales y naciones.

En suma, el intervalo de tiempo transcurrido entre el florecimiento de la aritmética política y el darwinismo vio un aumento en la incidencia de los datos y

su tratamiento científico en contextos cada vez más variados. El florecimiento de la aritmética política marca el primer antecedente de la Estadística como campo de conocimiento destinado al tratamiento de datos; no obstante, la tendencia al uso de datos se popularizó en las ciencias sociales, la epidemiología se constituye como campo de conocimiento al verse la necesidad de emplearlos en el trabajo clínico, y la industria se valió de estos para perfeccionar el control de calidad en sus procesos. Finalmente, es la probabilidad quien amplía el panorama del estudio de los datos al colocar sobre la mesa cuestiones como la detección de regularidades y el análisis de la variabilidad.

2.5 Datos y Guerras Mundiales: las Dos Caras de la Moneda

El siglo XIX fue el momento en el que se consolidó la extensión de las aplicaciones de las ciencias naturales a las ciencias sociales y biológicas (Yáñez, 2000), con lo que aumentó la tendencia al tratamiento de datos, favorecida por el uso cada vez más frecuente de la teoría de probabilidades. Sin embargo, la aplicación de los trabajos de Galton tuvo un capítulo aparte en el desarrollo de hitos históricos que marcaron para siempre el mundo, principalmente con la consolidación de los proyectos eugenésicos. Afortunadamente, la historia terminó mostrando que el procesamiento de datos a escalas sin precedentes resultó ser positivo para la humanidad porque fue la clave para finalizar más rápidamente la Segunda Guerra Mundial.

Las implicaciones de los trabajos de Galton, inicialmente orientados al estudio de variables antropométricas, no solamente incluían la recolección de datos sino su posterior modelización matemática a partir del uso de la distribución

normal. Con estos, se fundamentaron las bases de la eugenesia, el estudio de los mecanismos para el perfeccionamiento de la especie humana, favoreciendo la evolución natural (Saade, 2004).

La eugenesia, rápidamente difundida en Europa en la segunda mitad del siglo XIX, fue aceptada en Estados Unidos a comienzos del siglo XX por los problemas sociales y de inmigración experimentados en esa época. La tesis más importante de la eugenesia es la heredabilidad de los caracteres, esto es, que los rasgos hereditarios eran inmunes a los cambios de entorno, por lo que se podían transmitir de una generación a otra independiente del ambiente (Castro, 2014). Bajo esta premisa, se justificó promover el apareamiento entre personas con un buen material genético para preservar la calidad de la raza (eugenesia positiva) y esterilizar personas con material genético defectuoso (eugenesia negativa).

A partir de la recolección de datos y observación empírica se pudo detectar que el alcoholismo, la prostitución o la locura se heredaban del mismo modo que patologías como la hemofilia, el daltonismo o la ceguera, y que constituía una amenaza para el progreso, por lo que personas con estas características debían ser segregadas o eliminadas (Villela y Linares, 2011).

Si bien, la implementación de proyectos eugenésicos ayudó a la promoción de hábitos saludables de alimentación, sexualidad y cuidado físico, también justificó la persecución a grupos con características indeseables para la sociedad de la época. Malas interpretaciones de lo que se entendió por eugenesia llevaron a extremos con consecuencias fatales para la humanidad. Por ejemplo, el programa de eugenesia nazi ejecutado durante la Segunda Guerra Mundial para

purificar su raza resultó en la erradicación de comunidades estimadas por ellos como inferiores y la esterilización de alrededor de 350 mil personas con discapacidad, distorsiones mentales y homosexuales, entre otros (Proctor, 1992).

Una de las estrategias de los alemanes en la Segunda Guerra Mundial era la comunicación militar por medio de la Máquina Enigma, un dispositivo electromecánico indescifrable para la época y que les brindó ventajas en el conflicto bélico. De apariencia similar a la de una máquina de escribir convencional, funcionaba con un conjunto de rotores interconectados, cuyas posiciones se cambiaban periódicamente para confundir a cualquiera que quisiera interceptar sus mensajes. (BBC, 2020). Descifrar su funcionamiento, por parte de los británicos, era la clave para dar por terminada la Segunda Guerra Mundial.

La dificultad de esta labor estaba en que Enigma era capaz de configurar alrededor de 150 millones de millones de millones de diferentes mensajes, una empresa que se antojaba imposible (Sánchez, 2013). El conocimiento necesario para el éxito en el descifrado de Enigma residió en la criptografía, una combinación de técnicas matemáticas (permutaciones y teoría de grupos) y Estadísticas (análisis de frecuencias) para la interpretación de mensajes codificados.

El primer paso para lograrlo lo consiguió el matemático polaco Marian Rejewski, quien, a partir de sus conocimientos sobre permutaciones, información suministrada por el servicio de escucha del ejército polaco y agentes secretos franceses, logró reconstruir el cableado interno de Enigma, responsable de la generación de las configuraciones de Enigma. (BBVA Openmind, 2017). Quien

mejor aprovechó lo hecho por Rejewski fue Alan Turing, un criptógrafo reclutado por los británicos, que hizo parte de un equipo conformado por 10000 personas entre criptógrafos, matemáticos, ingenieros, lingüistas, jugadores de ajedrez, y otros, para interpretar las comunicaciones alemanas en el marco de la Operación Ultra (Fundación Telefónica, 2018). Motivado por la idea de crear máquinas autónomas al hombre, propuso replicar a Enigma inventando a Bombe, un dispositivo electromecánico capaz de buscar la configuración de los rotores de Enigma. La replicación de estas máquinas, aspecto en el que el matemático Gordon Welchmann fue clave, hizo posible salvar vidas.

Uno de los méritos de lo hecho por el equipo en el que Rejewski, Welchmann y Turing radica, entre otras cosas, en haber superado el listón en cuanto a capacidad de procesamiento de datos en la humanidad. Posiblemente, el antecedente más próximo se ubicaba con las máquinas tabuladoras de Hollerith hacia 1890, capaces de analizar datos de aproximadamente 60 millones de estadounidenses con motivo de los censos en Estados Unidos en tres meses, algo que se estimaba duraría cerca de 13 años sin la existencia de ellas (Ruiza, Fernández y Tamaro, 2004; Donoho, 2015).

Como se puede notar, lo mostrado con los proyectos eugenésicos y el fin de la Segunda Guerra Mundial es un ejemplo de cómo el uso de los datos puede resultar devastador para la humanidad o puede contribuir a su avance significativo. La ausencia de una adecuada reflexión sobre las implicaciones éticas de su uso y basarse en la clasificación de los datos para clasificar una persona como buena o mala, según sus características genéticas, hicieron de la eugenesia una práctica cuestionable en la comunidad científica. Por otro lado, la exhibición de generación

y procesamiento de datos en el marco de la Segunda Guerra Mundial fue una muestra del potencial beneficioso que pueden tener para la humanidad, pues permitió acortar la duración de la Segunda Guerra Mundial al menos dos años.

2.6 big-data: un Recubrimiento Abierto para la Humanidad

Lo hecho por Turing marcó un antecedente del caudal de datos que estaba por inundar prácticamente toda actividad humana al punto de llegar a usar el término big-data para dimensionar la cantidad de datos generados por la humanidad. Así mismo, los avances tecnológicos ocurridos en la segunda mitad del siglo XX causados por la importancia que disciplinas como la inteligencia artificial y la informática adquirieron en la vida de las personas, junto con eventos que demostraron la preparación de la humanidad para generar y almacenar datos a escalas cada vez más grandes, jalonaron diversas explosiones en la cantidad de datos al punto de llegar a un flujo cada vez más difícil de acotar.

Retomando a Turing, Bombe fue no solo la respuesta a una necesidad de índole militar sino la primicia de que era posible diseñar máquinas capaces de ejecutar y cambiar de tareas en forma autónoma, algo que no era posible en los tiempos de la Segunda Guerra (Preciado-Vidal, 2019). Un seguimiento a su obra permite constatar que sus esfuerzos iniciales estaban destinados a desarrollar máquinas capaces de trabajar en forma autónoma al ser humano, con lo que se gestan nuevos campos de conocimiento como la inteligencia artificial. En forma paralela a esto, llegó a conceptualizar la Máquina de Turing, con la cual se formalizó la idea de algoritmo y las bases de la teoría de la computabilidad (Alfonseca, 2000).

Así mismo, la construcción de máquinas capaces de simular el cerebro humano y la formulación, hacia 1950, de la prueba de Turing sobre la inteligencia de un ordenador, es el inicio de la inteligencia artificial (Harnad, 1992). De acuerdo con Warwick y Shah (2017) este consiste en una reformulación a la pregunta de si los computadores pueden pensar o no, con lo que se abría la posibilidad que las máquinas desarrollaran emular el pensamiento humano al menos de manera parcial.

La comunidad académica vinculada con la inteligencia artificial trabajó en diversos problemas de investigación que enriquecieron esta disciplina, entre estos, el uso que las máquinas hacían de los datos. Según Berlanga (2016) en las décadas del 50 al 80, el enfoque dominante en la comunidad de investigadores vinculados era el simbólico, y la cuestión a investigar era la transición de los datos al conocimiento por medio de técnicas de aprendizaje automático, en otras palabras, la construcción de algoritmos capaces de extraer conocimiento a partir de conjuntos considerables de datos, en términos de tamaño. Aunque en ese momento no se contaba con la capacidad para almacenar grandes cantidades de datos a bajo costo, la humanidad tardaría menos tiempo de lo que se esperaba para estar preparada.

Romero (2018) y Preciado-Vidal (2019) mencionaron algunos hitos que sucedieron a partir de la década de los 40 del siglo pasado y contribuyeron a que la humanidad pudiera gestionar grandes cantidades de datos. La explosión de información, gracias a un aumento en la demanda de investigaciones y publicaciones, ocasionó el crecimiento de las bibliotecas que se estimó se duplicarían cada 16 años según estimaciones de Fremont Rider en 1944. Así

mismo, la Teoría Matemática de la Comunicación desarrollada por Claude Shannon y Warren Weaver facilitó el trabajo con grandes volúmenes de datos en las décadas posteriores, porque fue respuesta a cuestiones como la compresión máxima de los datos y la tasa máxima de transmisión en el estudio de los procesos de transferencia de información en los sistemas de comunicación (Holik, 2016).

Continuando con Romero (2018) y Preciado-Vidal (2019), mencionaron la creación de memorias virtuales por Rudolf Grünsch en 1956 que abrió la posibilidad de procesar datos sin limitaciones de hardware. También, que en la década de los 60 se adquirieron habilidades en la captura de datos, donde merece mención el diseño de un sistema capaz de comprender la voz y los dígitos del 0 al 9 en presentado por William Dersch en 1962.

Adicionalmente, en el contexto de la automatización industrial, la llegada del primer controlador lógico programable optimizó el procesamiento de instrucciones. Con el surgimiento de las bases relacionales en la década de los 70 como modelo para acceder a grandes bases de datos compartidas, se encontraron formas de acceder a información sin saber su origen o cómo estaba estructurada (Llanos, 2010).

En la década de los 80 se llega a tal punto en la generación de datos que se llega a decir, parafraseando a la Ley de Parkinson, que “los datos se expanden hasta llenar el espacio disponible para el almacenamiento” (Preciado-Vidal, 2019, p. 10). Ante el potencial alcanzado hasta aquí, las técnicas de aprendizaje automático derivarían en lo que hoy se conoce como minería de datos, la cual ha

tomado lugar en los gobiernos, las empresas, el sector educativo, la minería de textos y el deporte, entre otras (Berlanga, 2016; Virseda y Román, s.f.)

La posibilidad de mejorar el diseño de computadores para que brinden mejores velocidades de procesamiento de datos, el desarrollo de la World Wide Web a partir de 1989 y la masificación del uso de Internet en la década de los 90, abrieron el campo de producción y consumo de información a cualquier persona que tuviera un computador, sin límites de tiempo o espacio geográfico, algo que intensificó la producción masiva de datos. Romero (2018) afirmó que se había estimado que la cantidad de información generada por la humanidad al final del siglo pasado era de $1.5 \cdot 10^{18}$ bytes.

El diseño y comercialización de teléfonos móviles cada vez más sofisticados han ampliado el espectro de la producción y difusión de datos a cada vez más millones de personas. Adicionalmente, la evolución de Internet ha hecho posible que los objetos, además de las personas, puedan estar conectadas entre sí, a tal punto que se habla de Internet de las Cosas.

Si se asume que, según el sitio Web Worldometers, en 2022 hay alrededor de 5500 millones de usuarios de Internet en el mundo y que en lo corrido del año en curso se han hecho cerca de $2 \cdot 10^{12}$ búsquedas en Google, el big-data es un recubrimiento abierto para la humanidad; si antes de Turing, la humanidad era capaz de contener a los datos que era capaz de generar, ahora es el big-data el que parece contener la humanidad.

2.7 Big-Data: Mucho más que big-data

Ante la mención de sucesos que dieron lugar al aumento de datos que se está presentando en el mundo, no cabe duda de que el papel del dato se ha vuelto cada vez más importante. Un estudio más profundo de las implicaciones que deja esta explosión llevaría a concluir que la influencia de los datos no solo está marcada por una cuestión de cantidad (big-data) sino por otras características que vuelven más complejo el asunto de manejar los datos a favor del desarrollo de la humanidad. La comunidad académica en general ha acuñado el término Big-Data para caracterizar las implicaciones que el big-data tiene en el mundo.

Existen diversas fuentes que tratan de ubicar el primer momento en el que aparece el término Big-Data. Desde el punto de vista de la capacidad de la humanidad para recolectar datos, Valero (2018) señaló que algunas fuentes ubican el inicio del Big-Data desde el paleolítico, hacia 180000 a. C., debido al interés de los grupos humanos de recolectar, contar y guardar datos. Sin embargo, en Preciado-Vidal (2019) se mencionó a Turing como el pionero del Big-Data, a raíz de sus trabajos para descifrar Enigma; quizá pensando en la gran cantidad de información que, para la época, tuvo que interpretarse.

Schweidel (2015) especuló que la primera referencia al término Big-Data se encontró en 1989, en un artículo de Erik Larson en la revista *Harper's Magazine* cuando había usado el término “keepers of big data” para sugerir que, en un futuro no muy lejano, quienes tuvieran acceso a sus datos en un futuro próximo seguirán de cerca sus preferencias de consumo y usarán los rastros de sus datos para generar publicidad sobre ellas. En el campo de la visualización científica, Michael

Cox y David Ellsworth usaron por primera vez el término Big-Data afirmando que “el campo de la visualización (científica) provee un reto interesante para los sistemas de computación: los conjuntos de datos son generalmente tan grandes que agotan las capacidades de la memoria principal, el disco local e incluso el disco remoto” (Cox & Ellsworth, 1997, p. 1).

Da la sensación, desde el mismo nombre del término, que el Big-Data hace referencia a conjuntos lo suficientemente grandes de datos como para desbordar las capacidades de almacenamiento existentes. Y no es para menos; en diversas fuentes se ha advertido que la cantidad de datos producidos por las personas, ya sean relevantes o no para algún propósito, oscilaba en el orden de 10^{21} bytes, (Pantula, 2011; Hoerl, Snee y DeVaux, 2014; Lateralía, 2016), lo que supera las capacidades de almacenamiento de los ordenadores tradicionales. No obstante, esto es apenas la primera característica asociada al fenómeno.

Según Cañón et. al. (2017) se debe distinguir al Big-Data del big-data, siendo este último el término que hace referencia a grandes cantidades de datos. Aunque no haya una definición genérica sobre Big-Data, hay un consenso para usarlo cuando se quiere hablar de datos generados y procesados por la sociedad en grandes cantidades (big-data), a velocidades cada vez más mayores, con grandes dosis de incertidumbre y en una amplia variedad de formatos (Camargo-Vega, Camargo-Ortega y Joyanes-Aguilar, 2015; Paredes-Moreno, 2015; Vásquez, 2020).

La primera característica del Big-Data es lo que hasta aquí se ha llamado big-data. La segunda es la exuberante velocidad a la que se generan datos, esto

es, un flujo cada vez más acelerado en su dinámica de entrada y salida a través de diversos canales. Paredes-Moreno (2015) señaló que diariamente se generaban cerca de 1.5 trillones de bytes de datos y, citando un informe de Oracle en 2012, que el 90% de los datos del planeta se habían generado en dos años. Petrelli (2018), en referencia a la generación de datos en la Industria 4.0, mencionó que en 2010 la cantidad de dispositivos conectados era alrededor de 10 billones y que en 2020 tal cantidad se había triplicado. Así mismo, Tabares y Hernández (2014) estimaban que la cantidad de datos digitales entre 2007 y 2020 aumentaría 44 veces.

La tercera característica del Big-Data, la incertidumbre, está asociada con aspectos como la veracidad, confiabilidad o calidad con la que se generan los datos. Joyanes-Aguilar (2013) mencionó que, según IBM, solo la tercera parte de los líderes de negocio no confía en la información que utilizan para tomar decisiones, aspecto que siembra dudas sobre el modo de actuar frente a ella. Tabares y Hernández (2014) mencionaron que la validez - ¿qué tan apropiados son los datos para el uso que quiero brindarles? - y la volatilidad - ¿cuánto tiempo pueden ser almacenados y por cuánto tiempo pueden ser válidos? - son retos igualmente importantes para generar valor en la información generada a través del procesamiento de los datos. Paredes-Moreno (2015) admitió que, inclusive los mejores métodos de limpieza de datos no erradican la imprevisibilidad inherente de algunos tipos de datos. Abordar la incertidumbre y reconocerla, con el fin de asegurar la calidad de los datos tratados, es un reto cada vez más complicado de asumir dado el volumen y la veracidad con la que llegan, pero abordarla es clave para extraer conocimiento de los datos que se presentan.

La cuarta característica del Big-Data, la variedad, está relacionada con la versatilidad de formatos en los que se generan los datos. A grandes rasgos, puede hablarse de dos grandes grupos de datos: estructurados, y no estructurados. De acuerdo con Cañón et. al (2017) los datos estructurados responden a fines específicos, son fáciles de encontrar, organizar e identificar y tienen una jerarquía rígida (Atomicidad, Consistencia, Aislamiento, Durabilidad), mientras que los no estructurados escapan a tal estructura y son producto principalmente de la actividad que se genera en la Web, documentos, imágenes, audios, entre otra clase de archivos. La irrupción de las redes sociales y demás actividades en la Web que implican generación de contenido digital han hecho que la presencia de datos no estructurados esté aumentando, al punto que estos ocupen el 90% de la cantidad de datos generados en el mundo. La organización y extracción de conocimiento de datos no estructurados ha representado uno de los retos más complejos para quienes trabajan técnicas de Big-Data.

Las referencias al fenómeno del Big-Data no solo hacen relación a las características de los datos que se producen hoy, también a las técnicas para tratarlos. En ese sentido, alrededor del fenómeno se ha hecho promoción del uso de sistemas tecnológicos con software de libre uso que facilitan la gestión de datos (Hadoop, MapReduce, HBase, Casandra, Mahout, entre otros). Escobar y Mercado (2019) señalaron que una caracterización del Big-Data obedece también a las etapas que comprenden la gestión de datos: la recolección de datos en la Web, su transformación en información, y la generación de valor agregado a partir de la información procesada.

Cañón et. al. (2017) se refieren al Big-Data como “un ecosistema creado por la emergencia concomitante de las ‘3C’ de Big Data, es decir (1) “Crumbs” (migajas); (2) “Capacities” (capacidades); y (3) “Communities” (comunidades)” (p. 6). La primera C se refiere a migajas digitales o el big-data, grandes cantidades de datos -estructurados o no- que reflejan el comportamiento y creencias de las personas, generados a partir de la interacción con dispositivos digitales (datos de escape), el uso de redes sociales y las conexiones con sensores físicos o remotos. La segunda C es en otras palabras lo que se conoce como Big-Data Analytics, el conjunto de herramientas, métodos y habilidades necesarias para el procesamiento y análisis de datos. La tercera C se refiere a los actores del ecosistema, desde los generadores de datos hasta los que consumen la información procesada; actualmente, son los llamados científicos de datos los que concentran la mayor participación en el ecosistema, pero aquí también caben los gobiernos, las empresas y personas en su doble postura de productores y consumidores de datos.

Por lo anterior, pensar que el ecosistema del Big-Data será el mundo mismo no parece lejano a la realidad. Valiéndose de sus características y la promesa de valor agregado ante la información obtenida por el procesamiento de datos, los gobiernos están adoptando estrategias vinculadas al Big-Data a sus ejercicios de gestión y las empresas se han vuelto más conscientes del potencial de los datos para la toma de decisiones. En esa misma medida, cada vez más personas, más allá de sus condiciones geográficas, nivel de escolaridad o sistemas de creencias, se comportan como generadores y consumidores de datos.

No es descabellado pensar que la sociedad misma se ha convertido en el hábitat del ecosistema Big-Data; en otras palabras: es la Sociedad del Dato.

3. La Sociedad del Dato y la Educación Estadística

¿Cuál es el significado de todo, Mr. Holmes?” “Ah, no tengo datos, no se lo puedo decir,” contestó. (Arthur Conan Doyle The Adventure of the Copper Beeches, 1892)

Si se sugiere que la sociedad puede llegar a ser el hábitat del Big-Data hace pensar que se ha vuelto en un flujo de datos constante, que se antoja inagotable. La sociedad del dato ha encontrado en ellos el potencial suficiente como para brindar mayor descubrimiento de sí misma y sus dinámicas. Quizás a esto se referían Mayer-Schönberger y Cukier (2013) cuando mencionaron que al incrementar la escala (o disponibilidad) de datos se abre camino a cosas que no se podían hacer a escalas más pequeñas.

La sociedad del dato descansa en el Big-Data como elemento configurador y al dataísmo como su orientación confesional. De hecho, con Big-Data se ha transformado la manera de concebir campos de conocimiento como la economía y la política, pero cada vez ha ganado más adeptos en otras esferas del ámbito productivo, al punto que la humanidad se ha visto tentada a aceptarlos como objeto de culto. Se puede admitir que los datos han tomado el sitio que, en épocas anteriores lo había ocupado la información o el conocimiento; cómo procesarlos y generar valor agregado a partir de ellos se ha vuelto el centro de la sociedad (Martínez y Soto, 2019).

No obstante, así como en los tiempos de los proyectos eugenésicos, no siempre el uso de los datos trae consecuencias favorables para la humanidad. De hecho, con el desarrollo de máquinas alimentadas con suficiente cantidad de datos y capaces de simular cada vez mejor al cerebro humano, se levantan cuestiones sobre el valor del hombre en un mundo cada vez más permeado por los datos. Desde un punto de vista teleológico, se pone en discusión cuál es el propósito del hombre en el marco de la sociedad del dato, poniendo en entredicho su valor frente a las máquinas que él mismo ha diseñado. Adicionalmente, la privacidad se ha convertido en un asunto sensible en un mundo donde la información fluye sin restricciones de tiempo o espacio.

Ante este panorama, la sociedad del dato como una manera de concebir el mundo en que vivimos genera implicaciones para nuestra existencia que habrá que revisar con cuidado. En este capítulo, se explicará cómo la sociedad del dato se ha gestado, revisando las interacciones de diversos campos de conocimiento que le han dado forma. Así mismo, se revisan rasgos de su identidad a través del sistema de creencias que maneja, cómo han permeado las actividades humanas y reformulado el propósito del hombre en la sociedad. Finalmente, se examina cuál sería la identidad del hombre en el marco de la sociedad del dato, y sus implicaciones en la formación que todo ciudadano debe tener con relación al análisis de datos, las cuales recaen en el dominio de la Educación Estadística.

3.1 ¿Y Cómo Llegamos a Esto?

La matemática es la ciencia que ayudó a consolidar disciplinas que dan forma a la sociedad del dato. La teoría de probabilidades ayuda a consolidar la

Estadística como ciencia y la herramienta que hace viable la detección de patrones, casos atípicos y correlaciones para extraer conocimiento de ellos, algo más emparentado con el campo emergente de la Ciencia de Datos; mientras que la lógica matemática y la física son las bases de la cibernética. La sociedad del dato es el resultado de la sinergia entre esta última, la informática y la neurociencia, siendo el cerebro humano el objeto de estudio por excelencia.

El desarrollo de la informática ha desembocado en el diseño de máquinas con conexiones similares a las del cerebro humano capaces de suplir sus limitaciones para procesar datos a escalas cada vez mayores. Para Harari (2015), uno de los aspectos que mejor explica el desarrollo de la humanidad es el aumento en la eficiencia en el procesamiento de datos caracterizado por: (a) un incremento en el número y variedad de procesadores, explicado en parte por las explosiones demográficas, y (b) sus conexiones entre sí, lo que facilita la libertad con la que fluyen los datos.

A manera de ejemplo, la capacidad de procesamiento de las máquinas tabuladoras de Hollerith, usadas en la segunda mitad del siglo XIX para agilizar los datos de los censos en Estados Unidos, rondaba los cientos de millones de datos (10^8); no obstante, para la Segunda Guerra Mundial, Alan Turing logró replicar la máquina Enigma que generaba alrededor de 10^{16} claves distintas. Se estima que, en la actualidad, la humanidad debe estar preparada para procesar tamaños de alrededor de 10^{21} bytes de datos.

El diseño de máquinas capaces de generar y procesar datos como los del Big-Data ha sido posible gracias al intento de simular el funcionamiento del

cerebro humano, siendo la cibernética la que ha aportado una mayor comprensión sobre este y las relaciones hombre-máquina. Como conocimiento científico desarrollado posteriormente a la Segunda Guerra Mundial, un antecedente epistemológico de la cibernética radica en Leibniz y su desarrollo de un *calculus ratiocinator*, raíz del cálculo lógico, con lo que se pudo considerar la idea que el pensamiento lógico pueda residir en el interior de una máquina (Soto, 2013).

Según Bowker (1993), los cibernéticos consideraban que las máquinas habían llegado a ser lo suficientemente complejas y su relación con el hombre tan intensa, que era necesario dotarla de un lenguaje común, rompiendo toda barrera comunicacional entre ellas y echando por tierra toda diferenciación entre lo humano y lo no humano. La cibernética resuelve este asunto bajo un esquema alternativo al método científico, centrándose en el estudio de los cambios que sufren los seres animados y artificiales con relación a su ambiente, y llegando a la conclusión que son equiparables en términos de la información y sus comportamientos. En palabras de Siles (2007), para la cibernética “todo ser, ya sea biológico, artificial o mecánico, puede definirse por la naturaleza de los intercambios de información que sostenga con su ambiente: todos son entes informacionales” (p. 89).

La neurociencia ha tomado de la analogía entre el cerebro humano y el funcionamiento de las máquinas para optimizar los procesos de aprendizaje en humanos, y gracias a la inteligencia artificial, en la forma como aprenden las máquinas. Con la idea de la Máquina de Turing se ha mejorado la capacidad para obtener algoritmos más complejos hasta el punto de que, en cierto modo, las máquinas tengan el potencial de superar su creador.

Cabañes y Rubio (2013) mencionaron que las redes neuronales y artificiales simulan el comportamiento de las redes neuronales biológicas cuando aprenden por sí mismas a reconocer patrones mediante un entrenamiento basado en el análisis automático y sistemático de algunos ejemplos, con lo que ganan la capacidad de aprender de acuerdo con su propia experiencia. También sugirieron que, algunas de ellas ya hacen parte de nuestros cuerpos, en algunos casos ayudando a suplir discapacidades y en otras generándolas, llamando la atención sobre las implicaciones sociológicas y antropológicas que esto pueda tener.

Haraway (1995) afirmó que una de las rupturas que marcaría la cultura científica en el siglo XXI sería la de hombre-máquina, en el sentido que el dualismo propuesto entre ellas terminaría por difuminarse. Aunque la programación inicial de las máquinas depende de la mano humana y su capacidad para procesar datos se basa en el razonamiento humano, han ganado autonomía en sus procesos de aprendizaje, llegando incluso a corregir sus errores.

Si bien, lo anterior supone un gran avance para la cibernética y la neurociencia, esto pone en juego consideraciones que amenazan con redefinir el propósito mismo de la existencia del hombre y su valor dentro de la sociedad. El lenguaje cibernético destrozó cualquier brecha entre hombre y máquina, dejando en entredicho el valor del hombre. Fabris, Bartolommei y Datteri (2008) hicieron alusión a Wiener (1998b) cuando llamaba la atención acerca de las implicaciones de la relación hombre-máquina:

Hoy nosotros somos los creadores, y los robots son nuestros golem. Pero ¿somos realmente como Dios? Y más aún, ¿cuál es la relación que

podemos tener con nuestras criaturas? (Mejor dicho, con nuestras creaciones) ¿Qué responsabilidad tenemos en estos casos?, ¿qué significa, en estos casos, tener la responsabilidad? (p. 263).

Tomando en cuenta que el mundo mismo se está convirtiendo en el ecosistema del Big-Data, está cobrando especial importancia procesarlos y adquirir conocimiento sobre ellos. En términos de lo anterior, da la sensación de que la responsabilidad humana es alimentar las máquinas con datos brutos y programarlas inicialmente. El credo de la sociedad del dato se basa en que todo el universo es un flujo incesante de datos: muéstrale a una máquina los datos que generas y te dirá quién eres.

3.2 Dataísmo: el Credo de la Sociedad del Dato

El vínculo entre el hombre y el dato se ha tornado cada vez más religioso, o al menos esa es la premisa del dataísmo, término con el que se identifica a la tendencia a considerar a los datos como objeto de culto. De acuerdo con Harari (2015) el dataísmo es resultado de la confluencia entre el darwinismo y las formulaciones de Turing gracias a la matemática; el darwinismo declara que los seres humanos son algoritmos bioquímicos, con la máquina de Turing se abrió espacio para cualificar la producción de algoritmos electrónicos, y la matemática es el nexo entre ambos campos de conocimiento para afirmar que las leyes matemáticas son aplicables por igual a seres humanos y máquinas.

La promesa que vende esta especie de religión descansa sobre la fidelidad con la que los datos pueden revelar más acerca de nosotros mismos. Según esta

filosofía emergente, los datos son una lente transparente para representar la historia del ser humano, todo lo que acontece con él y lo que pasará en el futuro; con lo que se conduce a una especie de “segunda Figura” donde el conocimiento y el progreso es movido a través de los datos (Han, 2014).

3.2.1 Las Facetas del Dataísmo

La euforia que desata la promesa del dataísmo ha permeado especialmente a los gobiernos, el manejo de la economía y la política; diversas voces han acuñado diferentes términos para describirlo como fenómeno. Se habla de *datificación* para referirse la tendencia a convertir la mayoría de los fenómenos cotidianos en datos, causado por la multiplicidad de fuentes de datos (humanas o artificiales) y la continuidad con la que los generan (Mayer-Schönberger y Cukier, 2013; Roca y Letouzé, 2016; Cañón et. al, 2017). El término *revolución de los datos* ha hecho parte de manifiestos en discursos relacionados con el desarrollo sostenible para llamar la atención sobre la calidad con la que se recogen datos para hacer más efectivas las rendiciones de cuentas y tomar decisiones (ONU, 2015).

La Organización para la Cooperación y el Desarrollo Económico (OCDE) también ha advertido del impacto que tiene no tener acceso completo a los datos o déficit en su recolección. Han mencionado que, de acuerdo con el Fondo para las Naciones Unidas para la Infancia (UNICEF), el 62% de los nacimientos ocurridos en los países menos desarrollados no quedan registrados (OCDE, 2015), lo que dificulta garantizar su derecho a la salud y la educación, así como posibilitar su formación como ciudadano del siglo XXI. Por esta razón, ha hecho un llamado a

las naciones en vías de desarrollo a guiar una revolución mundial de los datos dirigida por los países, esto es, “proporcionar los datos correctos a las personas adecuadas en el momento apropiado y con el formato preciso” (2015, p. 16).

Así mismo, con miras a contribuir al cumplimiento de los Objetivos de Desarrollo Sostenible propuestos por Programa de las Naciones Unidas para el Desarrollo (PNUD) para el 2030, la OCDE ha propuesto crear un índice de capacidad Estadística nacional y controlar su comportamiento, reconociendo el carácter público tanto de los datos nacionales y que “para aprovecharlos al máximo, los datos utilizados para entender, diseñar y seguir el proceso de desarrollo deberían actualizarse y divulgarse lo más ampliamente posible.” (OCDE, 2015, p. 17).

En el ámbito económico y gubernamental trabajar con datos parece inevitable. Cañón et al. (2017) hizo referencia a publicaciones que anunciaban el comienzo de algo que denominan *la era de los datos*. Destacaron a la revista británica *The Economist* donde en al menos un par de veces han hecho referencia al potencial de los datos; hacia 2010 en el artículo *The Data Deluge*, habían admitido que las empresas, los gobiernos y la sociedad estaban empezando a aprovechar el potencial de los datos, y en 2017 cuando se refieren a ellos como el nuevo petróleo, la nueva fuente de riqueza por el poder que se esconde en su extracción y explotación.

En WIRED (2014) se ha mencionado que la economía es digital, porque los datos son la clave del progreso y el adecuado funcionamiento de todo, desde el gobierno hasta las compañías locales. Lohr (2015) mencionó que a diferencia de

la sociedad industrial donde el carbón, el hierro o la gasolina fueron los combustibles del desarrollo, son los datos la fuente vital de la economía; así mismo, sugirió que la clave de los individuos y las organizaciones para ser competitivos radica en su habilidad para explotar, proteger y manejar los datos.

La política brinda ejemplos del potencial escondido detrás de los datos. La campaña de Barack Obama utilizó en 2012 técnicas de Big-Data para sacar mayor ventaja a sus rivales por la presidencia de los Estados Unidos, y la firma Cambridge Analytica, que en el marco de las elecciones estadounidenses de 2016 obtuvo los datos de 50 millones de usuarios de Facebook para analizar sus rasgos y manipular sus preferencias, en las elecciones que adjudicaron a Donald Trump como sucesor de Obama. Estos testimonios han abierto la necesidad de regular en los países el uso de datos y controlar el manejo que tienen las entidades que accedan a ellos.

El carácter divino con el que se ha dotado a los datos no es exclusivo de los gobiernos o empresas pues en otros campos de conocimiento han comenzado a sacar provecho de los datos para diversos fines. Por ejemplo, Lohr (2015) muestra un buen uso de los datos: el del Hospital Emory, donde existen máquinas que son capaces de manejar 160.000 datos por segundo, cuantificando los cuidados intensivos necesarios para las personas enfermas. Otro caso es la astronomía; los telescopios se convirtieron en generadores de grandes cantidades de datos, lo que ha obligado a los astrónomos a trabajar en forma interdisciplinar con ingenieros informáticos para saber cómo manejarlos.

Los dataístas sugieren que el dato proporciona un lenguaje común a las personas y disciplinas, con lo que las barreras entre los campos de conocimiento se rompen, algo compatible con Wiener (1998a) quien consideraba la comunicación como el elemento común de todas las disciplinas. A manera de ejemplo, desde una visión dataísta, la quinta sinfonía de Beethoven, la burbuja de la bolsa y el virus de la gripe son patrones de flujo de datos que pueden analizarse usando conceptos y herramientas similares, con ello que músicos, corredores de bolsa y biólogos moleculares puedan comprenderse entre sí (Harari, 2015). Con esto, se abren como nunca en la historia oportunidades para la generación de conocimiento y el surgimiento de nuevas disciplinas como resultado de la interdisciplinariedad.

La presencia del dato en ámbitos cada vez más variados lleva a pensar que estamos ante una *revolución digital*, o, en otros términos, la Cuarta Revolución Industrial. Klaus Schwab, fundador del Foro Económico Mundial, en un artículo del 2016 la caracterizó como un fenómeno basado en las transformaciones derivadas de la Tercera Revolución Industrial y una fusión de tecnologías que están borrando las fronteras de separación entre lo biológico, lo digital y lo físico.

Si bien Schwab no mencionó de forma explícita a los datos como el recurso central en esta nueva época, sí hace énfasis en la influencia de la inteligencia artificial, en los que la gran cantidad de datos ha sido un gran responsable de su avance como campo de conocimiento. Adicionalmente, hizo referencia a fenómenos relacionados con Big-Data que le han dado forma: multiplicidad de conexiones mediante dispositivos electrónicos, capacidades sin precedentes de almacenamiento e Internet de las Cosas, entre otras.

Sugiere que, a diferencia de las tres revoluciones anteriores, los cambios experimentados durante esta nueva era son exponenciales, en el sentido que las transformaciones son cada vez más veloces, de mayor alcance y de mayor impacto. Schwab (2016) afirmaba al respecto que:

La Cuarta Revolución Industrial, finalmente, no solo cambiará lo que hacemos sino quienes somos. Afectará nuestra identidad y asuntos relacionados con ella: la idea de privacidad, nuestros patrones de consumo, el tiempo que dedicamos a trabajar y el que nos queda libre, como desarrollamos nuestras carreras, cultivamos nuestras habilidades, conocemos personas y nutrimos nuestras relaciones. Ya está cambiando nuestra salud y llevándonos a un “quantified-self”, y más pronto de lo que pensamos puede llevarnos a una realidad humana aumentada. La lista es interminable porque está acotada solo por nuestra imaginación. (2016, párr. 26).

El dato, al ser un elemento cada vez más presente en toda actividad cotidiana, es el protagonista de la Cuarta Revolución Industrial. Es la nueva fuente de riqueza, el motor de la economía y una nueva forma de control, con un poder determinante en la toma de cualquier clase de decisiones; la falta de ellos detiene el progreso; controlar su volumen es un reto cada vez mayor para la humanidad, su comprensión tiene implicaciones en todos los órdenes de la existencia.

Hoerl, Snee & DeVaux (2014) afirman que “Big-Data está aquí para quedarse y está revolucionando la industria de los negocios, el gobierno, la ciencia, la educación, la medicina, y virtualmente a todos los aspectos de la

sociedad” (2014, p. 223). Como se puede apreciar, las cuestiones por saber cómo procesar y qué hacer con los datos, que empezaron como un problema exclusivo de los gobiernos, se ha trasladado a todos los aspectos de la sociedad, lo que hace pensar que el dato es actualmente el elemento central del desarrollo en la humanidad.

3.2.2 *¿Dato, Luego Existo?*

Atribuir características mesiánicas a los datos genera debates que colocan en entredicho consideraciones existenciales del ser humano. El dataísmo sostiene que el hombre es un algoritmo bioquímico, en otras palabras, un sistema de procesamiento de datos, con lo que queda en discusión su condición de superioridad frente a sus creaciones (máquinas) y seres vivientes.

En una entrevista al diario el País (Galindo, 29 agosto 2018), Harari señaló que poco a poco se está llegando a un punto en el que las máquinas adquieran mayor conocimiento sobre el hombre que él mismo: sus datos personales, sus gustos, los lugares que frecuenta y hasta las reacciones ante lo que consume - solo por mencionar algunas-. Esto también supone una dependencia del ser humano cada vez mayor al algoritmo y con ello la pérdida de capacidad de tomar decisiones por sí mismo. La dependencia del hombre hacia el uso del dato así sea involuntaria, se torna inevitable.

Oppenheimer (2018) sugiere que los ejercicios de procesamiento de datos podrían llegar a usarse cada vez más para sustituir valores como la experiencia y el buen juicio; lo cual es una muestra que la visión antropocéntrica (humanista) del universo quedaría comprometida, pues podría llegar un punto en que las

máquinas logren no solo ser independientes del hombre sino incluso llegar a superarlo. Si algo define el peso específico del hombre en la realidad del dataísmo es su contribución al flujo y procesamiento de datos.

En ese sentido, el valor de la experiencia humana no está en los detalles que la conforman sino en que sea compartida. No hacerlo la convierte en algo sin trascendencia, así como la aprobación (o en algunos casos, desaprobación) de otros usuarios que se unen al flujo de datos; en otras palabras, el *me gusta* es como el amén digital (Han, 2014) y las redes sociales son el templo donde se repite la experiencia de compartir datos.

Todas y cada una de tus palabras y actos forman parte del gran flujo de datos, que los algoritmos te observan constantemente y que les importa todo lo que haces y sientes [...] estar desconectado del flujo de datos supone arriesgarse a perder el sentido mismo de la vida. ¿Qué sentido tiene hacer o experimentar algo si nadie se entera y si no aporta algo al intercambio global de información? (Harari, 2015, p. 147).

Ya no importará por qué o para qué se hacen las cosas, lo importante es que se hacen, y que su comportamiento se puede seguir y medir transparentemente siempre y cuando haya suficientes datos (Han, 2014). Inclusive, se puede afirmar que la interacción de las personas con la Web no solo alimenta el flujo de datos actual, también esconde un correlato de lo que son. A manera de ejemplo, Youyou, Kosinski y Stillwell (2014) en el marco de un estudio psicológico para explicar los juicios de los participantes basados en sus perfiles de Facebook muestran que con 10 “me gusta” un modelo computacional puede

conocer mejor a una persona de lo que lo hace un colega de trabajo y bastan 300 “me gusta” para que lo conozca mejor que su propia pareja o a sí mismo.

Los rastros que las personas dejan en la web representados por los datos que se comparten representan problemas para la intimidad de las personas. Harari (2014) advirtió que cada vez son más personas las que están dispuestas a conceder su privacidad cuando publican todo lo que se hace en la red a costa de pertenecer al flujo de datos existente en la actualidad.

Desafortunadamente, las secuelas que dejan los datos en la red no se limitan solamente a huellas rastreables en la web; si las máquinas tienen suficiente conocimiento de nosotros y los modelos que se hagan de ellos tienen influencia directa en las decisiones que afectan la humanidad se espera que sean objetivos y sin sesgo, algo que no está pasando. Harari (2017) señala que la IA tiene el potencial de reducir el valor político y económico de la humanidad y que todo podría terminar en una dependencia de la red para tomar decisiones importantes sobre aspectos tan delicados como la salud y la supervivencia.

Gallardo (2020) en una entrevista a Julio Rojas con relación a la emergencia de nuevas tecnologías, advierte que “ahora se crean sobre la marcha las reglas del juego y las tecnologías se han convertido en un aliado valioso para resolver problemas sociales, pero también para incrementar los índices de inequidad en el mundo” (p. 24). O'Neill (2016), en su libro *Armas de Destrucción Matemática (ADM)*, se refiere así a los modelos matemáticos a partir de los datos, para sugerir que refuerzan problemas que larga data en el mundo. Señala ejemplos de cómo una aplicación inconsciente y poco ética del Big-Data puede

ocasionar daños irreparables en la vida de las personas: negar un préstamo, no poder empezar la universidad, no poder acceder a un trabajo o a servicios de salud, ser condenado injustamente a prisión, entre otros casos.

Al respecto, señala que “con su promesa de eficiencia y justicia, distorsionan la educación superior, acrecientan la deuda, incitan a las penas de prisión en masa, golpean a los pobres en prácticamente todas las coyunturas y socavan la democracia” (O’Neill, 2016, p. 313). Lejos de defender al Big-Data y de proponer un desmantelamiento gradual de las ADM, sugiere incorporar elementos éticos al modelamiento matemático, que sepa mantener un equilibrio entre la justicia y los beneficios, y que imite mejor el juicio ético humano. Así mismo, que quienes trabajan con datos adviertan al mundo de los posibles abusos y malentendidos de sus modelos.

Con todo y que contribuir al flujo de datos implique revelar demasiado de su intimidad, y que se dude de su carácter infalible y objetivo, el propósito del ser humano y su búsqueda por alcanzar algo superior a él parece reducirse a fusionarse con el flujo de datos en la red. Todo parece indicar que su misión sería alimentar máquinas superinteligentes y perfeccionar la inteligencia artificial. Si bien, el dataísmo pareciera ofrecer un camino para comprender el pasado y el futuro, este tipo de iluminación solo es un modo distinto de esclavitud.

Así las cosas, el hombre ya no viviría para sí, sino para ser parte del flujo de datos; no ser parte de esta dinámica implicaría perder, incluso, su papel en el mundo. En palabras de Harari:

Cuando los humanos perdamos nuestra importancia funcional para la red, descubriremos que, después de todo, no somos la cúspide de la creación. Las varas de medir que nosotros mismos hemos consagrado nos condenarán a sumarnos en el olvido a los mamuts y a los delfines fluviales chinos. En retrospectiva, la humanidad resultará ser sólo una onda en el flujo cósmico de datos. (2015, p. 429).

Si el hombre está destinado a ser una unidad de análisis dentro del flujo de datos y tomando en cuenta fenómenos como el Big-Data, da la sensación de que las máquinas pudieran sobrevivir a pesar de una probable extinción de los seres humanos, e incluso replicarse sin intermediarios humanos. Aunque las anteriores afirmaciones puedan tornarse demasiado dramáticas y requieren un cuidadoso escrutinio, queda claro que el potencial de los datos en la actualidad parece no tener techo.

Los datos ya no son simples registros propios en las redes informáticas, pues están adquiriendo un papel cada vez más relevante en la sociedad. Ya el collar electrónico del que hablaba Deleuze (Sibilia, 2006) no es un elemento externo al ser, hace parte de él. Las fronteras entre la máquina y el hombre son cada vez menos evidentes y su fusión parece ser el destino final de ambos, por lo que la Sociedad del Dato y sus implicaciones se convierten de vital importancia para la reflexión académica y los alcances que los datos puedan tener sobre nosotros.

3.3 El Prosumidor de Datos

En un mundo donde el dato es omnipresente, indefectiblemente la humanidad es prosumidora de datos. Por un lado, la necesidad de que las personas cuenten con datos de calidad sobre el mundo que les rodea, unido al uso de tablas y gráficos estadísticos para representar fenómenos de interés en medios de comunicación y redes sociales, ha convertido a las personas en consumidoras de datos; lo que ratifica lo advertido por Nikola Tesla hacia 1926, cuando dijo que “cualquier persona será capaz de acceder y analizar grandes cantidades de datos utilizando un dispositivo lo suficientemente pequeño que cabrá en el bolsillo.” (Ciampagna, 2015). Por otro lado, las implicaciones del desarrollo de la web 2.0 y el surgimiento de las redes sociales han facilitado que las personas se vuelvan productores de contenido, y, por tanto, de datos.

Ya que es inevitable escapar a este rol dicotómico, formar a las personas como consumidoras de datos permitirá que la persona amplíe su conocimiento del mundo que le rodea; que asuma posturas críticas ante la información que recibe sobre los datos que le rodean, y que tome decisiones más fundamentadas sobre asuntos que afectan su vida, ideas emparentadas con el concepto de Alfabetización Estadística. Por otro lado, formarlas como productoras de datos alentará el desarrollo de competencias deseables en todo profesional como lo son la investigación y la generación de conocimiento, ideas más compatibles con el desarrollo de Pensamiento Estadístico.

3.3.1 *El Consumidor de Datos y Alfabetización Estadística*

La cuestión por el acceso a los datos ha sido un tema de interés mundial, cuando menos, desde 1948, momento en que la Organización de las Naciones Unidas (ONU) reconoció la necesidad de hacer buenos ejercicios de recolección de datos y difusión de la información de las naciones; motivos suficientes para preocuparse por cualificar el personal que trabajaba en los departamentos nacionales de Estadística. El fin último siempre ha sido garantizar la calidad en los procesos de tal forma que cualquier ciudadano pudiera acceder a la información desprendida de los datos, para tomar decisiones que afectarían su estilo de vida; según la ONU, “datos mejores, vidas mejores” (ONU, 2015). Desde la ONU se reconoce que “necesitamos datos y Estadísticas locales para asegurar que todos los niños tengan acceso a la educación, y datos y Estadísticas mundiales para vigilar los efectos generales del cambio climático.” (ONU, 2015).

Si se habla de personas que no pertenecen a agencias o departamentos de Estadística entonces cabe mencionar a la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), amparada en el derecho a la libre expresión que incluye, entre otras cosas, investigar y recibir información, declara que la Alfabetización Informacional y Mediática (AMI) es indispensable para la formación de personas del siglo XXI, pues dota de competencias que le brindan al ciudadano los elementos para ejercer su derecho a la libre expresión.

La Alfabetización Informacional contempla aspectos del uso de la información como el acceso, la evaluación y su uso ético; mientras que la mediática incluye entender, evaluar y comprometerse con los diferentes medios de

comunicación de la información que se emite. Dentro de los diversos formatos en los que puede venir, no se descarta el uso de tablas, gráficos y estadísticos asociados de los datos que dan forma a los fenómenos que se representan. Obtener AMI será vital para lograr el cuarto objetivo de la agenda del PNUD ya mencionada anteriormente: lograr una educación de calidad en población joven y adulta.

Esta necesidad radica, entre otras cosas, en el tiempo y las razones que las personas gastan en medios web y redes sociales, sitios donde los datos y la información se expanden rápidamente. En un reporte sobre tendencias digitales a nivel mundial, Kemp (2022) concluyó que la media mundial de tiempo de consumo de internet al día en personas entre 16 y 64 años es de 6 horas y 58 minutos al día, y que en Colombia esta cifra es significativamente mayor (10 horas y 3 minutos). Así mismo, entre las razones principales para navegar por Internet se encuentra la necesidad de buscar información (61%) y estar al día con noticias y eventos (53.1%).

No es casual entonces encontrar que, por ejemplo, si una persona se declara interesada en política, interactúe con datos y Estadísticas propias de los fenómenos electorales o los índices de popularidad de una personalidad pública. También, es común encontrar Estadísticas de rendimiento en competencias deportivas durante la transmisión de eventos deportivos, lo cual enriquece la experiencia del consumidor al observar una competencia y los profesionales del deporte analizan e interpretan esta información para mejorar el desempeño. A la hora de la compra de un producto usando Internet o al bajar una aplicación a un teléfono celular, es inevitable encontrar opiniones que otros usuarios han emitido

sobre el producto en interés a través de diversos esquemas de calificación presentados en forma gráfica.

En medio de esta realidad, las personas pasan tiempo consumiendo información en diversos formatos, dentro de los cuales se incluye aquella que proviene de la Estadística. Entonces, cobra sentido lo que desde la Educación Estadística se ha mencionado sobre la importancia de formar competencias que permitan a las personas analizar e interpretar tanto los datos como la información Estadística que se usa en medios de comunicación o, la que surja del estudio de fenómenos de interés común; o, en otras palabras, volver la sociedad Estadísticamente culta.

Es así como surgió un primer concepto que incluye el análisis de datos y otras competencias adjuntas mencionadas por los investigadores, denominado Alfabetización Estadística –de aquí en adelante SL–. En Wallman (1993) se encontró la primera definición formal de SL: “habilidad de entender y evaluar críticamente resultados estadísticos que permea nuestras vidas a diario, junto con la habilidad de apreciar las contribuciones que el pensamiento estadístico puede hacer en las decisiones públicas, privadas, profesionales y personales.” (p.1). A partir de esto algunos investigadores añadieron elementos a la definición de Wallman y configuraron modelos para explicar la SL como constructo.

Por ejemplo, Watson (1997) señaló que se trata de comprender la información Estadística en función del contexto donde está ubicada y propuso una jerarquía de tres niveles para formar ciudadanos Estadísticamente cultos –básico o de comprensión Estadística; intermedio o de comprensión contextual; avanzado

o de posturas críticas—, en el que la habilidad para asumir posturas críticas sobre la información encontrada en medios de comunicación es la condición para catalogar a un ciudadano como Estadísticamente culto. Este modelo se refinó en Watson y Callingham (2003) con seis niveles, en coherencia con la taxonomía SOLO: idiosincrático, informal, inconsistente, consistente no crítico, crítico y crítico matemático.

Moore (1998) sugirió que la SL es el conjunto de ideas sobre Estadística que las personas que no son especialistas en Estadística requieren en el siglo XXI. Garfield (1999) y Snell (1999) comentan que la SL implica un entendimiento básico de lenguaje estadístico y la habilidad de encontrarle sentido a las construcciones que se hacen con Estadística (tablas, gráficos, estadísticos). Para Gal (2002) la SL es la interacción de dos habilidades a saber:

(a) interpretar y evaluar críticamente información Estadística, argumentos relacionados con los datos o los fenómenos estocásticos, los cuales pueden encontrarse en diversos contextos, y cuando sea relevante y (b) la habilidad de discutir y comunicar reacciones a tal información Estadística, esto es, entender el significado de tal información, opiniones sobre sus implicaciones, o las preocupaciones acerca de la aceptabilidad de las conclusiones dadas (p. 3).

Entre otras cosas, este autor señala que en la formación de SL intervienen aspectos como la literacidad, la Estadística, la matemática, el contexto y la habilidad crítica; además de las disposiciones, aspecto vinculado a las creencias, las actitudes y la capacidad para asumir posturas críticas.

Franklin et. al. (2005) en la primera versión de *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) plantean un documento marco con el ánimo de parametrizar los procesos de enseñanza y aprendizaje en Estadística y probabilidad en los Estados Unidos, y ha sido adoptado como referente mundial. De este se infirió que la SL era la meta de formación Estadística en los estudiantes de cualquier nivel educativo, sugiriendo como objetivos, formar capacidad para:

(a) formular preguntas orientadas a la recolección de datos, organizar y mostrar aspectos relevantes de ellos, (b) seleccionar y usar métodos apropiados para analizar datos, (c) desarrollar y evaluar inferencias y predicciones basadas en los datos y (d) entender y aplicar conceptos básicos de probabilidad. (p.5).

Budgett & Pfannkuch (2007) sugirieron que las actividades para el fomento de la SL debían llevar a los estudiantes a valorar críticamente la información contenida en medios de comunicación y construir informes usando estadísticos. A nivel iberoamericano el término se ha traducido por Batanero (2002; 2013) como cultura Estadística, debido a la necesidad identificada por la UNESCO de reconocer el rol del ciudadano en la actividad Estadística, así como su valor cultural.

Algunos investigadores afirmaron que los elementos característicos de la SL apuntan a la idea que los ciudadanos sean buenos consumidores de datos. Gould (2010) señaló que la dicotomía SL-ST es compatible con el doble rol consumidor-productor; Serradó Bayés (2013) citó a Schield (2010) cuando afirmó que la SL es “la habilidad de leer e interpretar resúmenes estadísticos en los

medios cotidianos: en gráficos, tablas, afirmaciones y ensayos.” (p. 23) y aclara que la SL es necesaria para quienes son consumidores de datos. Weiland (2017) advierte que las definiciones de Wallman (1993) y Gal (2002) apuntan a que las personas requieren ser efectivos consumidores de datos.

Otro aspecto que Weiland (2017) identificó como una palabra presente en las caracterizaciones de la SL es la palabra *crítica* y rediseñó el concepto para proponer un marco de trabajo que desarrolle una *Alfabetización Estadística Crítica* –de aquí en adelante, SLC– en la escuela. Para Weiland, la SL no solo implica consumir información Estadística usando conceptos básicos, sino usarla como un lente desde el que se puede adquirir una visión más amplia del mundo. Entre los elementos que caracterizan su construcción, destacó que la SLC capacita a las personas a comprender la influencia de los datos, las investigaciones Estadísticas y los factores sociopolíticos e históricos, en la construcción de los discursos que moldean la sociedad, y la manera como la gente consume información Estadística y emite sus opiniones sobre esta.

3.3.2 *El Productor de Datos y el Pensamiento Estadístico*

Las cuestiones sobre el procesamiento y producción de datos que parecían interés solamente de los científicos y los entes gubernamentales han alcanzado al ciudadano del común, más allá de su *modus vivendi* o condición socioeconómica. Aunque la mayoría de los datos que las personas generan no son estructurados como en las bases de datos clásicas y muchos de ellos no tienen aparente relevancia, los registros de sus actividades los convierten en productores de datos.

Un caso testimonial de lo anterior ocurre con el Self-Tracking. Según Lupton (2016), es una práctica adoptada por personas usada para conocer más de ellos mismos a partir de la observación y registro de sus actividades cotidianas, y el sentido que le otorguen a tales datos. El ejemplo más revelador de esta práctica es el Quantified-Self (QS), término que, según el mismo autor, apareció gracias a Gary Wolf y Kevin Kelly en una publicación de *Wired Magazine* en 2007, y que ha hecho carrera a partir de las comunidades de personas creadas alrededor del Self-Tracking en función de la calidad de vida.

Quienes participan de estos grupos se involucran en proyectos basados en ST, los cuales comprenden la generación de datos a partir de intereses propios, la observación de detalles relevantes, el razonamiento sobre sus datos y la interpretación de su propia información. Han (2014) mencionó que para hacer esto posible se dota al cuerpo de sensores que registran datos de forma automática logrando obtener datos sobre variables que caracterizan el estado de salud, el estado de ánimo, y (dicho de alguna forma) el rendimiento espiritual, entre otras cosas. La promesa de valor que el QS vende es que se puede alcanzar conocimiento de sí mismo –o al menos de cosas relevantes– por medio del análisis de los datos que se generen.

Un escenario más cercano de las personas como productoras de datos ocurre en las redes sociales. Citando nuevamente a Kemp (2022), el 58.4% de la población mundial está conectada al menos a una red social y un porcentaje importante de estos (21.4%) las usan principalmente para publicar cosas sobre su vida, con lo que la generación de datos no estructurados es cada vez mayor. Recordando que el 62.5% de la población mundial se conecta a Internet (ver

sección anterior), dentro de las razones principales para hacerlo también se encuentran la compra de artículos on line (58.4%), actividades de negocios (30.8%) y de networking (21.4%).

De las cifras anteriores, puede inferirse que existe una necesidad latente por ser parte del flujo de datos a cualquier costo, a tal punto que se ha convertido en necesidad el compartir toda clase de experiencia vivida. Tal parece que un eslogan de esta sociedad es, como diría Harari, “Si experimentas algo, regístralo. Si registras algo, súbelo. Si subes algo, compártelo” (2015, p. 420). Se tenga o no conciencia de esto, compartir nuestras experiencias involucra un ejercicio de producción de datos; no importa si se estudia o trabaja, si se crea o se copia, un solo segundo de vida es suficiente para generarlos. Sin importar que alguien tenga interés en la Estadística o se vea como un científico de datos, y aun a costa de poner en riesgo su intimidad e individualidad, el ser humano ha formulado su interés en ser parte del flujo de datos actual no solo como alguien que observa las experiencias de otros en su entorno, sino como alguien que contribuye a su sostenimiento y circulación.

Lo anterior lleva entonces a preguntarse, si basta con visualizar a las personas únicamente como consumidoras de datos, cuestión que recae en el ámbito de la Educación Estadística, campo en el que más investigación se ha desarrollado con relación a temas en torno al análisis de datos. En efecto, por algún tiempo se mantuvo la idea que la SL era una meta deseable de formación Estadística; no obstante, algunos de los investigadores mencionados reconocieron elementos relacionados con la producción de datos.

Por ejemplo, Watson (2006) en una nueva revisión a partir de su anterior investigación con Cullingham (2003) mencionó que un aspecto importante de la SL es reconocer cómo los datos son producidos. Weiland (2017) propuso la SLC pensando en elementos como el tratamiento de la variación y la diferencia, conceptos centrales en Estadística pero que requieren no ser solamente comprendidos sino entendidos en forma más amplia. También, la última versión de GAISE en 2016 considera factores como el aumento de estudiantes en cursos de Estadística a nivel mundial, el surgimiento de la Ciencia de los Datos como un campo emergente y la exigencia del uso de recursos tecnológicos, para revisar las expectativas en la formación Estadística que se habían propuesto una década atrás.

Por estas y más razones, acudieron a un nuevo constructo, el de Pensamiento Estadístico –de aquí en adelante ST–, más compatible con la idea de productor de datos (Brown y Kass, 2009; Gould, 2010), para designar la meta de formación en Estadística más importante. A diferencia del caso de la SL donde la Educación Estadística ha provisto más definiciones, en el caso de ST se ha intentado dar con cualidades con las que debe contar un pensador estadístico o alguien que piensa Estadísticamente.

Si bien la primera mención de ST se puede encontrar en una paráfrasis que Wilks (1951) hace del escritor británico Herbert George Wells al afirmar que “llegará el día en que el pensamiento estadístico será una condición tan necesaria para la convivencia eficiente como la capacidad de leer y escribir” (p. 5)”, ST y SL son conceptos contemporáneos. A pesar del abuso en el empleo de este constructo, en parte por malas traducciones de palabras como *thinking* o

reasoning al español, la primera definición se ubica en el contexto del control de calidad.

Se dice que fue Deming (1993) el pionero del ST gracias a lo declarado sobre el ciclo Planear, Hacer, Verificar y Actuar (PHVA) en el contexto del conocimiento profundo de un sistema. Sin embargo, es la American Society for Quality (ASQ) quien da la primera definición formal con base en los planteamientos de Deming. Según esta, ST es:

Una filosofía de aprendizaje (una forma de pensar, distante de las técnicas Estadísticas) y acción basada en tres principios (a) todo ocurre en un sistema de procesos interconectados, (b) la variación existe en todos los procesos y (c) comprender y reducir la variación es clave para el éxito. (ASQ, 1996, p.5)

En el contexto de la Educación Estadística, Moore (1997) afirmó que es un modo de razonamiento sobre los aspectos que involucran el trabajo con datos; sobre todo el reconocimiento de su necesidad en términos de la valoración de un contexto y los procesos que afectan su producción. Afirma que para enseñar ST se debe llevar a las personas a comprender la necesidad e importancia de producir datos y prepararlas para tratar su variabilidad, a través de la modelación.

Con relación a las cualidades que deben formarse para ayudar a las personas a pensar Estadísticamente, Wild, Pfannkuch y Triggs (1997) concluyeron que el ST involucra otras clases de pensamiento como el estratégico, el adaptativo y el necesario para la modelación. Wild y Pfannkuch (1999; 2000) mencionaron la transnumeración, entendida como el manejo de diversas formas de representar los

datos, el uso de la imaginación y del sentido común, y ponerse en el lugar de otros.

En la misma dirección que los autores anteriores, Chance (2002) afirmó que es ST implica una comprensión de porqué y cómo se hacen las investigaciones en Estadística, lo que incluye reconocer procesos que van desde la formulación de preguntas que guían el proceso de recolección de datos hasta los métodos que se eligen para probar hipótesis sobre los datos. En ese sentido, quien piensa Estadísticamente asume actitudes escépticas hacia la forma como los datos son obtenidos, relaciona la información Estadística hacia la obtención de datos sin quedarse solamente en tecnicismos, y piensa más allá de los textos guía, algo ratificado en Behar y Grima (2004). Hoerl & Snee (2010) mencionan que el pensador estadístico es cuidadoso en el análisis de la variabilidad presente en los datos, reafirmando a Moore (1997).

Álvarez Madrigal & López Ríos (2015) concluyeron que el ST debe ser formado en los cursos de Estadística para la formación de ingenieros. En medio de una revisión conceptual sobre ST, Riascos (2014) lo contextualiza como:

la capacidad mental, a partir de los conocimientos y la formación general en Estadística, que desarrolla un individuo al leer, analizar e interpretar datos, capacidad que se complementa con la actitud que este individuo asume al enfrentar y manipular dichos datos. (p. 63).

3.4 Otros Conceptos Relacionados con SL y ST

A partir de una revisión sobre investigaciones en Educación Estadística, Andrade, Fernández y Álvarez (2017) observaron la preeminencia de un enfoque

basado en la SL y ST, algo que parece ratificar las recomendaciones hechas en GAISE (2016) de perseguir como fin principal de la formación Estadística al ST sin descuidar el fomento de la SL. No obstante, las investigaciones muestran el uso de varios constructos relacionados con SL y ST, probablemente el más mencionado sea el Razonamiento Estadístico (RE). A partir de revisiones sobre este concepto, Riascos (2014) trató de definirlo como

el proceso que permite a los seres humanos extraer conclusiones a partir de premisas o hechos, evidenciados en datos, acaecidos previamente, ayudados por las técnicas y teorías Estadísticas disponibles para las personas en el contexto de su conocimiento y utilizados en el marco de la situación. (p. 17)

Riascos (2016) afirma que existe una necesidad por delimitar los términos expuestos anteriormente, algo que ratifica Andrade, Fernández y Álvarez (2017) cuando afirmaron que la SL, el ST y el RE pueden indagarse más. También, Riascos (2016) señaló que, según investigaciones reseñadas por delMas (2002), los conceptos SL, RE y ST suelen utilizarse indiferentemente en investigaciones, y que tampoco hay una definición estándar de ellos. Así mismo, en investigaciones latinas parece haber una tendencia a confundir RE y ST por errores de traducción al español de investigaciones anglosajonas.

A pesar de que aún no hay consensos universales sobre sus diferencias y similitudes, hay trabajos que intentaron hacerlo. Por ejemplo, Snee (1990) mencionó que el ST trasciende los textos escolares, a diferencia de los otros constructos, algo ratificado en Behar y Grima (2004). Ito (2000) trata de hacer una

distinción con base en los diversos niveles de formación, cuando mencionó que la SL (entendida por el autor como capacidad Estadística) debería ser objetivo para la formación obligatoria, media y cursos introductorios de Estadística; el ST debería ser meta para todos los niveles de educación, incluyendo futuros analistas de datos y educadores en Estadística, mientras que el RE es meta para futuros estadísticos en todos los niveles de formación.

delMas (2002) recogió los estudios de Rumsey (2002), Garfield (2002) y Chance (2002) para mostrar posibles relaciones entre los dominios SL, ST y RE. Si se visualizará a cada uno de ellos como eventos probabilísticos, un primer enfoque sostiene que son independientes dos a dos, pero no disjuntos; esto es, se puede elegir enfatizar en uno de ellos sin importar que tanto se trabajen los otros dos, y que incluso en el mejor de los casos pueden trabajarse los tres dominios. Entre tanto, un segundo enfoque argumenta que el RE y ST están contenidos en la SL; en otros términos, que la SL es la meta general de formación Estadística para toda persona, idea más compatible con la guía GAISE.

Desde lo anterior, delMas (2002) estableció diferencias en términos de la evaluación. Argumentó que evaluar SL implica habilidades para construir ejemplos que ilustren los conceptos centrales de Estadística, manejar en forma relacionada diversas formas de representación e interpretar resultados de un procedimiento estadístico; evaluar ST pone en juego acciones como juzgar, evaluar críticamente interpretaciones y preguntar; mientras que evaluar RE implica cuestionar el proceso de producción de datos. También, consideró que la SL es el elemento básico para la formación de RE y ST.

Siguiendo con el aspecto de la evaluación, en la guía *Assessment Resource Tools for Improving Statistical Thinking* (ARTIST, 2006) se proveen recursos para los docentes a fin de entender más claramente estos términos; de acuerdo con esta y en términos de la Taxonomía de Bloom, el ST está relacionado con los tres niveles más altos (análisis, síntesis y evaluación), el RE con el segundo nivel (comprensión) y la SL con el primer nivel (conocimiento).

Otros conceptos que han aparecido con relación a la SL y el ST se han derivado de los trabajos de Rumsey (2002) cuando propuso los términos *competencia y ciudadanía Estadísticas* para hablar de las metas de formación Estadística. Para Rumsey la competencia Estadística es como el conocimiento básico para el fomento de ST y RE, término que también trabajó Schield (2010) cuando definió que “es la capacidad para producir, analizar y resumir Estadísticas detalladas en encuestas y estudios” (p.2). Entre tanto, Rumsey (2002) definió *ciudadanía Estadística* como una meta que trasciende del ST y RE, algo ratificado en estudios hechos por Nikiforidou, Lekka y Pange (2010) sobre tendencias de SL en educación superior, cuando sugirieron que la ciudadanía Estadística es el lugar de convergencia entre SL, ST y RE.

En Batanero et. al. (2013), tratando de establecer analogías con la acepción *sentido numérico* encontrada en el campo de la Educación Matemática, definen el *sentido estadístico* como la unión entre cultura Estadística (traducción hecha por ella del término *statistical literacy* o SL) y el RE.

3.5 Estadística y Matemática: ¿Estadísticamente Independientes?

Si bien, parece haber un consenso que la SL y el ST son metas que traducen los objetivos de formación en Estadística, y, por tanto, en análisis de datos, es necesario constatar si, en medio de un contexto como el que ofrece la Sociedad del Dato, las personas están siendo preparadas en función de tales metas. Para verificarlo, se debe revisar cuál es el papel que la Estadística ocupa en los currículos y en qué modo la Educación Estadística ha evolucionado como campo de investigación.

Uno de los reclamos más frecuentes desde las investigaciones en Educación Estadística es la necesidad de reconocer a la Estadística como un campo particular de conocimiento con problemáticas e intereses diferentes al de la Matemática; por lo tanto, requiere de una práctica didáctica y pedagógica que se distinga de la Matemática. Batanero (2013) citando a Moore (1991), menciona que

La Estadística es una disciplina científica autónoma, que tiene sus métodos específicos de razonamiento. Apoya esta afirmación en el hecho de que la Estadística no ha surgido de la matemática, sino de una serie de ciencias que se han apoyado en la matemática. Más aún, la relación entre Estadística y matemáticas no es biunívoca; la Estadística toma conceptos matemáticos para el desarrollo de sus métodos, en cambio la matemática no usa conceptos estadísticos. La Estadística tiene también controversias específicas (por ejemplo, sobre el significado de la probabilidad), y la posición que un estadístico toma sobre ellas tiene un impacto inmediato en su práctica. (p. 2).

Burrill y Biehler (2011) en medio de un estudio sobre currículos en Estadística en varios países, notaron que las ideas Estadísticas más importantes que se desprenden de ellos se resumen en conceptos como: gráficos, distribución, correlación y asociación, probabilidad, muestreo, inferencia, variabilidad aleatoria y datos. González-Gómez (2014), en una revisión sobre aspectos diferenciadores entre la Estadística y la Matemática, menciona el tratamiento de la variabilidad y los datos.

La noción de variabilidad, a menudo emparentada con el concepto de incertidumbre, azar o aleatoriedad, ocupa un lugar preponderante en Estadística, algo que no ocurre en la Matemática. Moore (1997) menciona que la variabilidad es omnipresente en todo proceso de producción de datos y que la Estadística proporciona herramientas para tratar con ella. Adicionalmente, para una persona que quiera analizar datos o un estadístico, el conocimiento de la variabilidad es un aspecto central en Estadística, pues está presente en todo intento de modelización Estadística y es un rasgo característico de quienes aprenden a pensar Estadísticamente (ASQ, 1996; Hoerl & Snee, 2010).

Si se quisiera hacer una distinción del dato en Matemática y el dato en Estadística, también hay diferencias significativas. Mientras que en Matemática no es relevante el modo como los datos son obtenidos e incluso pueden existir sin la necesidad de un contexto, en Estadística los datos se trabajan en función del contexto que los rodea y su estudio implica abordar ideas como su producción, sus representaciones (tabulares o gráficas), la forma, la centralidad, la dispersión, y cuestiones relacionadas con la posibilidad que nos brindan de hacer inferencias.

Moore (1997) hace una distinción entre las implicaciones que tiene la forma como se trabaja el dato en Matemática y Estadística en las siguientes líneas:

En Matemáticas, donde el contexto aplicado es mucho menos importante, los ejemplos improvisados a menudo funcionan, y los profesores de matemáticas se vuelven hábiles para inventar ejemplos en el acto (¿Se requiere una función para ilustrar la regla de la cadena? No hay problema: simplemente se diseña). En Estadística, por el contrario, los ejemplos improvisados no funcionan porque no proporcionan una auténtica interacción entre el patrón que los forma y el contexto. (p. 803).

Batanero et. al. (2013) complementaron lo anterior explicando que, en Matemática, el contexto donde se ubica el dato es imaginario, en Estadística el contexto es mucho más relevante que el tratamiento teórico del concepto.

Si bien la Estadística ha recogido elementos de la Matemática para su desarrollo como campo de conocimiento, no se puede decir que dependa de la Matemática. Como se ha podido apreciar, la Estadística cuenta con una serie de inquietudes ajenas a las de la Matemática, las cuales pueden visualizarse en el abordaje de conceptos centrales como la variabilidad y el dato, entre otras ideas más. Precisamente, el tratamiento de conjuntos de datos es la diferencia más notoria entre ambas disciplinas, debido al valor que el contexto tiene para sus futuras interpretaciones y análisis.

3.5.1 Educación Estadística y Educación Matemática: ¿Estadísticamente Independientes?

La Educación Estadística empezó a ser investigada a mediados del siglo XX gracias a la creación del Comité de Educación del ISI, motivado por recomendaciones de la UNESCO de impulsar su desarrollo. Vere-Jones (1997) aclaró que a pesar de que el ISI fue creado en 1885, fue con el Comité de Estadística que empezó una organización sistemática apuntando a realizar recomendaciones para enseñar Estadística; para tal fin, viene realizando intervenciones en los procesos de enseñanza, publicaciones y organización de eventos académicos. También, que hasta 1976 el énfasis del ISI estuvo en formar personal especializado en el tratamiento estadístico de cuestiones gubernamentales. Luego de 1976, concentró esfuerzos para promover la Educación Estadística en colegios y universidades.

A pesar de esto, el interés en este campo ha sido desigual en el mundo si se comparan los países desarrollados con los latinos debido a que la incursión de la Estadística en el currículo fue más tardía en estos últimos (Martínez, 2014). En medio de una revisión de tesis durante el período 2000-2014, Andrade, Álvarez y Escobar (2017) muestran que cerca de la mitad de ellas provienen de Estados Unidos y que en el concierto iberoamericano es España el que más aporta.

De Garfield & Ben-Zvi (2008), Del Pino & Estrella (2012) y Groth (2015) se puede inferir que la Educación Estadística no ha logrado todavía separarse por completo un campo separado de la Educación Matemática, lo cual puede explicarse porque el lugar que se ha asignado para la Estadística está dentro de la

Matemática. Adicionalmente, quienes se han encargado de alimentar la investigación en Educación Estadística vienen de diversos campos de conocimiento (educadores matemáticos, psicólogos, educadores e investigadores en tecnología educativa) pero no principalmente de la Estadística.

Por ello, Garfield y Ben-Zvi (2008) mencionaron que “para la gente interesada en leer sobre estas publicaciones, la Educación Estadística puede parecer una disciplina invisible y fragmentada” (p. 21). Investigadores en Educación Estadística han reclamado sobre las diferencias entre la Estadística y la Matemática, y que no reconocerlas es un obstáculo en la formación Estadística.

3.5.2 Enseñanza de Estadística: ¿Otro Curso de Matemáticas?

A pesar de lo advertido en las anteriores secciones sobre las diferencias entre la Estadística y la Matemática, y que los intereses de investigación entre la Educación Estadística y la Educación Matemática son distintos, aún persiste el hecho que se enseñan ideas Estadísticas como si fuesen conceptos matemáticos. Zapata (2011) afirmó que

Por años, la enseñanza de la Estadística se ha desarrollado al amparo de la enseñanza de las matemáticas y por ser las matemáticas una ciencia de naturaleza determinística, la enseñanza de la Estadística también ha adoptado este carácter. La Estadística en sí misma ostenta una naturaleza no determinística porque la variación es una de sus particularidades. Parece ser que, a pesar de esta peculiaridad, por mucho tiempo, la enseñanza de la Estadística ha ignorado su naturaleza. (p. 236).

González-Gómez (2014) añadió que la enseñanza de la Estadística fundamentada en el aprendizaje de algoritmos “es insuficiente para que el profesor se pueda sentir en capacidad de ofrecer una clase de Estadística que privilegie el pensamiento aleatorio y la toma de decisiones.” (p. 20). También, Donoho (2015) sugirió que es imposible enseñar a analizar datos si estos se enseñan como se hace en Matemática.

De Campos (2016) se infiere que ni la SL ni la ST pueden ser enseñadas con instrucción directa desde el aula y solo pueden ser observadas cuando se evidencian ciertas actitudes relacionadas con el uso de los datos. Zapata (2016), en una investigación que da cuenta sobre la promoción del ST en el aula, sugirió que este se fomenta muy poco pues las investigaciones evidencian una tendencia a utilizar datos que otros han producido, enseñar conceptos en forma descontextualizada y sin una intención investigativa.

De MacKay (2016) y Pino et. al (2017) puede inferirse que no hay referencias sobre investigaciones que evalúen los niveles de SL ni ST en adultos, lo cual establece un panorama incierto que no ayuda a establecer si lo que se ha investigado con relación a estos constructos ha tenido efectos o no en la formación Estadística de ellos. Weiland (2017), en el contexto de la educación escolar, mencionó que así se invoque a la SL como parte del currículo en matemáticas, en el mejor de los casos es abordada superficialmente, dando prioridad al manejo algorítmico.

Lo anterior no ofrece un panorama alentador sobre la enseñanza de la Estadística y los resultados de la formación que se está brindando en escuelas y

universidades, ya que sus conceptos centrales se están ofreciendo desde un punto de vista determinista y algorítmico. La Estadística se está presentando como una extensión de un curso de matemáticas básicas; y aunque aplicar un recetario de fórmulas quizá garantice aprobar Estadística, esta acción dista mucho de los objetivos reales de formación Estadística y, por tanto, de los resultados esperados en función de la sociedad del dato.

3.5.3 *El Profesor de Estadística: ¿Preparado para Enseñar?*

Tal vez, uno de los escenarios donde mejor se refleja la falta de distinción entre la Estadística y la Matemática, sea en el profesor que enseña Estadística. Ya sea que se hable de la escuela o de la universidad, el personal que asume el reto proviene de campos diferentes a la Estadística: en su mayoría licenciados en Matemática o profesionales de otras áreas.

Aunque en general, quienes se responsabilizan de enseñar Estadística en las universidades son matemáticos o profesores de matemáticas, dependiendo de quien ofrezca los cursos de Estadística en las universidades pueden existir una amplia variedad de perfiles. Pinto et. al. (2017) mencionaron que, en los niveles secundario y universitario, cerca del 60% no tienen una formación específica en enseñanza de la Estadística, y que es usual encontrar profesionales de áreas como los ingenieros, físicos, biólogos, entre otros. González (2019) ratifica de algún modo lo anterior, mostrando que cerca del 70% de los profesores de Estadística de una institución privada no tienen formación específica en el área.

Si bien no está documentado, existe la tendencia de que en las facultades de las universidades prefieren que sus profesores de Estadística sean

profesionales afines con sus perfiles de egreso, mientras que en los departamentos de ciencias básicas prefieren profesionales del área de las ciencias (matemáticos, físicos, etc.). Aun así, todos ellos asumen el reto de enseñar Estadística sin importar que se visualicen como educadores en Estadística o no, gracias a la idea que el profesor de Matemática, el profesional con fuerte formación cuantitativa, o alguno que haya tomado algún curso de Estadística Descriptiva, Inferencial o Probabilidad puede asumir cursos de Estadística sin problema; detrás de esto pueden existir razones de carácter administrativo para realizar esta clase de asignaciones académicas; o que existe poca oferta de docentes altamente cualificados para enseñar Estadística.

Así las cosas, quien enseña Estadística asumen el reto con la consigna de preparar y/o reaprender sobre la marcha, los conceptos relacionados en un parcelador, y para esto se vale de recursos como libros impresos o digitales, bases de datos ficticias o software especializado; pero no es consciente de las implicaciones didácticas de su decisión porque enseña Estadística del modo como lo haría en Matemática.

En cualquier caso, Rocha (2012) diagnosticó que en las facultades y los departamentos de ciencias básicas hay pocas evidencias de ejercicios reflexivos acerca de cómo se enseña Estadística. También advierte de otros inconvenientes; entre los más críticos se encuentran la falta de uso de la Estadística de los profesores y estudiantes pues “un porcentaje muy bajo de los profesores admiten poseer bases sólidas y actualizadas en métodos y técnicas Estadísticas” (p. 35).

Adicionalmente, mencionó del poco desarrollo de herramientas para la enseñanza de la Estadística y la poca oferta de programas de formación y actualización en Estadística. Ratificando lo anterior, en un documento de la *American Statistical Education* (ASA) encaminado a dar recomendaciones a profesores que enseñan Estadística en Estados Unidos, se señaló que “debido a la pobre oferta de programas o cursos especializados, los docentes no tienen la culpa por su falta de su falta de preparación para enseñar conceptos estadísticos efectivamente” (Franklin, et. al, 2016, p. 49).

3.5.4 ¿Cómo son los Resultados Finales de Aprendizaje en Estadística en Educación Superior?

El referente nacional para evaluar la calidad de la educación superior a nivel universitario son las pruebas SABER-PRO, las cuales se aplican a todo estudiante que haya cumplido por lo menos el 75% de sus créditos de su programa de pregrado. Dentro de los módulos evaluados aparece el *razonamiento cuantitativo*, entendido como “el conjunto de elementos de las matemáticas (sean estos conocimientos o competencias) que permiten a un ciudadano tomar parte activa e informada en el contexto social, cultural, político, administrativo, económico, educativo y laboral” (ICFES, 2017a, p. 31).

El reporte de resultados del ICFES en 2021 muestra que el 59% de los estudiantes universitarios que presentaron la prueba a nivel nacional en 2020 están en los niveles más bajos de razonamiento cuantitativo y que la media nacional –en una escala de 1 a 300– fue de 152 puntos, la cual no tiene diferencias significativas si se le compara con los resultados desde 2016. Añaden

también que los profesionales de la educación resultaron con el peor rendimiento (134 puntos) entre las áreas mencionadas por el reporte. De acuerdo con la definición de los niveles de desempeño esto implica que la mayoría tiene dificultades para identificar información implícita en representaciones gráficas no usuales y establecer la validez de procedimientos por medio del uso de procedimientos aritméticos, algebraicos, variacionales y aleatorios; lo que le complica llegar a resolver problemas usando herramientas cuantitativas.

Otro módulo dentro de SABER-PRO, dirigido a un público más reducido, es el de pensamiento científico en matemáticas y Estadística, entendido como “la capacidad que tienen los estudiantes de comprender, analizar y afrontar situaciones reales o abstractas con rigor científico.” (ICFES, 2017b, p.16) y tradicionalmente elegido por pregrados en ingeniería, ciencias naturales y exactas. Según ICFES (2020a) el 62% de estudiantes de universidades privadas y el 53% de públicas que presentaron la prueba SABER-PRO de 2019 obtuvo resultados en los niveles más bajos del módulo de razonamiento cuantitativo. También, desde ICFES (2020b) se puede inferir que la situación es aún más crítica cuando se consideran los puntajes del módulo específico razonamiento científico, en el que el porcentaje de estudiantes con niveles bajos de desempeño son 84% en privadas y 79% en públicas (ICFES, 2020b).

En el contexto de la evaluación en pregrados de ingeniería colombianos, la Asociación Colombiana de Facultades de Ingeniería (ACOFI) ha aplicado anualmente, desde 2006, el Examen de Ciencias Básicas (EXIM) con el fin de tener un conocimiento global de las fortalezas y debilidades de los futuros ingenieros colombianos. EXIM contiene cuatro marcos conceptuales, dentro de los

cuales está el marco matemático; que a su vez incluye cinco dominios conceptuales, entre ellos el aleatorio “que incluye el estudio de situaciones relacionadas con formalizaciones matemáticas de evaluación poblacional en términos probabilísticos y de aplicación a múltiples situaciones.” (ACOFI, s.f., p. 8).

Los últimos resultados disponibles se encuentran en 2021 y están presentados por marcos conceptuales. La media de los puntajes en el componente aleatorio fue de 50.8 (en una escala de 0 a 100); no obstante, el informe de ACOFI (2022) sobre el marco conceptual matemático señaló con relación al componente aleatorio de la prueba que muestran dificultad para aplicar conceptos como las medidas de tendencia central y las distribuciones de probabilidad.

Desde luego, la preocupación por el desempeño en estas pruebas es tema de debate en las instituciones de educación superior y lo anterior ratifica los resultados de una formación Estadística que en gran parte de las universidades sigue haciéndose al amparo de la Matemática. Da la sensación de que no existe evidencia alguna que las pruebas descritas anteriormente involucren conceptos estadísticos o manejo de datos evalúan SL o ST, y que urge tener instrumentos que permitan ver si lo investigado en Educación Estadística ha surtido o no algún efecto. Habrá que revisar si al interior de las facultades, la formación Estadística de los estudiantes universitarios es atendida con cuidado.

3.6 Ciencia de Datos: ¿Otro Espacio para el Análisis de Datos?

Probablemente, el área de conocimiento más asociada con el análisis de datos sea la Estadística. Al revisar un buen número de definiciones de la palabra

Estadística se encuentran una serie de ideas que confirman dicha intuición; en efecto, Wu (s.f.) recolectó significados puestos en diccionarios, que coinciden en que es la ciencia de la colección y el análisis de datos numéricos. Fienberg (2014), en un ejercicio similar, pero revisando investigaciones en Educación Estadística mencionó las siguientes:

- “Conjunto de conceptos, reglas y métodos para (1) recolectar datos, (2) analizar datos, y (3) obtener conclusiones de estos” (Iversen y Gergen, 1997, p. 4).
- “Estadística es la ciencia y el arte de recolectar, analizar y hacer inferencias de los datos” (Mosteller et. al., 1961, p. 2).
- “Estadística es el arte de aprender de los datos. Está relacionado con recolectar datos, su subsecuente descripción y su análisis, el cual guía a menudo a establecer conclusiones” (Ross, 1996, p. 5).

En definiciones encontradas en textos escolares, universitarios o glosarios se menciona que:

- “Es la ciencia que comprende una serie de métodos y procedimientos destinados a la recopilación, tabulación, procesamiento, análisis y presentación de datos cuantitativos y cualitativos” (INEI, 2006).
- “La Estadística estudia los métodos científicos para recoger, organizar, resumir y analizar datos, así como para sacar conclusiones válidas y tomar decisiones razonables basadas con tal análisis.” (Universidad Rómulo Gallegos, 2013).

- “El arte y la ciencia de recolectar, analizar, presentar e interpretar datos.” (Anderson, Sweeney, Williams, Camm y Cochran, 2019, p. 3).

No todas las definiciones mencionan en forma explícita al análisis de datos.

En algunas de ellas, se menciona que la Estadística es:

- “El arte de hacer conjeturas matemáticas sobre preguntas desconcertantes.” (Freedman et al. 1978, p. xiii).
- “La forma de razonar, a través de una colección de herramientas y métodos diseñados para ayudarnos a entender el mundo.” (De Veaux et al, 2006, p.2).
- “Es la ciencia de los datos” (Moore, 1991, citado por Batanero, 2013, p. 3).

La última definición citada anteriormente menciona un campo que ha surgido como producto del acelerado crecimiento en la producción de datos manifestado en el fenómeno del Big-Data y que nutre a la Sociedad del Dato: la Ciencia de los Datos. Probablemente, la mención hacia esta nueva disciplina surgió a raíz del llamado que hizo John Wilder Tukey, en su obra *The Future of Data Analysis* (1962), a reflexionar sobre cuál tenía que ser el campo de acción de su trabajo como estadístico. Tomando como punto de partida un contexto en el que el desarrollo de técnicas inferenciales ocupaba la mayor parte de la labor de los estadísticos estaba dejando al análisis de datos en un plano secundario. Donoho (2015), citando párrafos introductorios de la obra de Tukey, menciona que:

Por mucho tiempo he pensado que era un estadístico interesado en inferencias de lo particular a lo general. Pero he visto la Estadística Matemática evolucionar, lo cual me ha causado asombro y dudas... en resumen, siento que mi interés central está en el análisis de datos, en el cual incluyo, entre otras cosas: procedimientos para analizar datos, técnicas para interpretar los resultados de tales procedimientos, formas de planear la recolección de datos para hacer de este análisis algo más sencillo, más preciso o más adecuado, y toda la maquinaria de resultados estadísticos (matemáticos) aplicables para analizar datos. (2015, p.10).

Lo que Tukey estaba impulsando era un crecimiento de la Estadística como disciplina, algo que pudiera llamarse *Análisis de Datos*. Como consecuencia, la Estadística resultó incrustada en la Ciencia de Datos, nombre más llamativo que el sugerido por Tukey, y nombrado por William S. Cleveland, en su trabajo *Data Science: An Action Plan for Expanding the Technical Areas of the field of Statistics* (2001). Los científicos de datos, apoyados en la obra de Tukey, sugieren que la Ciencia de Datos es movilizada por cuatro fuerzas: (a) las teorías formales de la Estadística, (b) el desarrollo acelerado de los computadores y dispositivos electrónicos, (c) el reto de contar con grandes cantidades de datos en una diversidad cada vez más amplia de campos de conocimiento, y (d) el énfasis de la cuantificación en una variedad cada vez más amplia de disciplinas.

La secuela que dejó el llamado de Tukey terminó siendo que gran parte de la comunidad académica en Estadística se volviera disidente del trabajo estadístico para erigirse como científicos de datos. Dentro de la revisión de Donoho (2015) sobre la evolución de la Ciencia de Datos, se encuentran

comentarios en los que se retaba a hacer Ciencia de Datos sin Estadística, o declarar que es la parte menos importante de la Ciencia de Datos.

De todos modos, en el campo de la Educación Estadística, asumir la existencia de este campo emergente ha generado opiniones orientadas a reconocer su existencia y tomarla como parte del reto del trabajo con datos. Por ejemplo, Gould (2010) reconoció que la forma como son producidos los datos en la actualidad no es la misma que en épocas anteriores, debido a la presencia de datos no estructurados que ocupan gran parte de la generación de datos de la humanidad, con lo que llamó la atención de revisar los métodos estadísticos del momento para analizarlos y extraer información de ellos, algo en lo que la Ciencia de Datos podría ayudar.

Blanco (2018), mencionó diversos autores que reconocen que la Estadística actual es muy distinta a la que se enseñaba a finales del siglo pasado; en parte por la emergencia de la Ciencia de Datos, traducida en la influencia que ha tenido en la proliferación de nuevos métodos estadísticos (Utts y Horton, 2016); la presencia de grandes datos (Cope y Kalantzis, 2016; Mayer-Schönberger y Cukier, 2013; Science, 2011) y los retos que ha supuesto para el procesamiento de información y el aprendizaje a partir de los datos (ASA, 2015). No obstante, hay voces que no aceptan la Ciencia de Datos del todo. Fienberg (2014, citando a Broman, 2013) señaló que “si estás analizando datos, estás haciendo Estadística. Puedes llamarlo Ciencia de Datos o informática o analítica o cualquier cosa, pero aún sigue siendo Estadística” (Fienberg, 2014, p. 5).

Aunque de Donoho (2015) se puede inferir que, en el contexto estadounidense, la Ciencia de Datos es una disciplina que lleva al menos cinco décadas de desarrollo, no sucede de la misma manera en países como Colombia; prueba de ellos es que, de acuerdo con la revisión de Medina (2021), solo el 4% de los artículos en Ciencia de Datos está escrito en español. Por otro lado, aunque se intuye que la formación básica de un científico de datos incluye la Estadística, la computación y la Matemática, la cuestión por la enseñanza de la Ciencia de Datos o elementos cercanos a ella no aparece aún en el radar de la comunidad científica, ni tampoco qué elementos puede brindar la Ciencia de Datos en la construcción de una formación basada en el análisis de datos.

Por otro lado, al observar lo que en Ciencia de Datos se entiende por análisis de datos, no hay tantas definiciones que lo reconozcan como parte de su objeto de estudio. El ejemplo más representativo quizá se encuentre en el código de conducta del científico de datos emitido por la Data Science Association (s.f.) donde se menciona que Estadística “significa la práctica o ciencia de recolectar y analizar datos numéricos en grandes cantidades” (parr. 15) mientras que la Ciencia de Datos es “el estudio científico de la creación, validación y transformación de los datos para crear significado” (parr. 1). En otra definición que Medina (2021) brindó, citando a IBM, se observó que:

La Ciencia de Datos es un campo interdisciplinario sobre procesos y sistemas para obtener conocimientos o perspectivas de grandes volúmenes de datos de diversas formas, ya sean estructurados o no estructurados, lo que implica la continuación de algunos de los campos de análisis de datos

como la minería de datos y el análisis predictivo, así como el descubrimiento del conocimiento y la minería de datos. (p. 5).

Al parecer, el enfoque de la Ciencia de Datos está en la extracción de conocimiento a partir de los datos, mientras que el de la Estadística se encuentra más en los procesos de recolección y análisis. Así mismo, se nota una preocupación por aclarar que el entorno de oficio de quienes trabajan con datos, sobre todo en el campo de la Ciencia de Datos, es el Big-Data; grandes tamaños, velocidades y variedades de datos.

En ese sentido, haría falta encontrar cuáles son las explicaciones que, tanto la comunidad Estadística como la de Ciencia de Datos brindan con relación al análisis de datos: ¿qué lugar otorgan al análisis de datos en la formación de profesionales que trabajan con datos?, ¿es el análisis de datos una acción directamente relacionada con el estadístico y con el científico de datos?

4. Metodología

*“La Teoría Fundamentada no es una teoría,
sino una metodología para descubrir teorías
que dormitan en los datos” (Strauss, 2004, p. 51)*

En lo que sigue se da cuenta de las técnicas utilizadas en esta investigación para alcanzar los objetivos que se han propuesto. Se explican las razones que llevaron a elegir el enfoque cualitativo en detrimento del cuantitativo. Se ofrece una breve descripción de la Teoría Fundamentada como la base del diseño metodológico y cómo se adaptaron las diversas etapas de su aplicación en la obtención de resultados. Finalmente, se realiza un breve análisis descriptivo de la población de estudio, esto es, los Programas de Pregrado en Estadística y Ciencia de Datos (PECD) en Colombia.

4.1 Enfoque de la Investigación

Partiendo de las necesidades implicadas en el objetivo general de esta investigación, es necesario un enfoque que visibilice la manera como los sujetos – las universidades– perciben el objeto de estudio –el análisis de datos–. Al recordar la revisión presentada en el anterior capítulo, se manifiesta la necesidad de brindar explicaciones sobre el análisis de datos, lo cual es posible construyendo teoría a partir del levantamiento y revisión documental. Dicho esto, obtener estadísticos descriptivos o hacer análisis inferenciales basados en muestras representativas, no es funcional para tal propósito; se requiere el tratamiento de información presentada en textos completos o fragmentos de ellos, los cuales se analizan con

técnicas alejadas de lo que se hace usualmente en Estadística. Por esta razón, se asumirá un enfoque cualitativo, que según Martínez (2004) surge como respuesta a necesidades propias de las ciencias sociales que no pueden ser llenadas del todo por los métodos cuantitativos.

En esta investigación, se parte del supuesto que el objeto de estudio está inmerso en un contexto social, histórico y educativo que debe ser visibilizado y comprendido. Así mismo, se busca comprender lo que las universidades oferentes de programas de pregrado en Estadística o Ciencia de Datos –de aquí en adelante PECD– manifiestan con relación a la emergente Sociedad del dato, el profesional que prometen en sus ofertas y el lugar que le otorgan al análisis de datos en sus propuestas formativas. En ese sentido, un enfoque cualitativo resulta conveniente porque trata de identificar la profundidad de las realidades a investigar, su estructura dinámica, manifestaciones y razones de sus comportamientos; lo que implica mirar las cosas desde una perspectiva sistémica y dialéctica (Martínez, 2004).

En esta investigación se asume que los sujetos son universidades y que la información que se analizará documentalmente proviene de sus micrositios web, donde es posible encontrar documentos institucionales, brochures, fragmentos de textos o narraciones extensas que describen el modo como estas perciben el ejercicio de la educación superior en general. Según Denzin y Lincoln (2018) la investigación cualitativa se vale de una gran variedad de material empírico que refleja momentos rutinarios y problemáticos en las vidas de las personas y aprovecha este acopio de información para descubrir el significado que le otorgan a las acciones que realizan a diario.

Cabe precisar que la construcción teórica a realizar no contempla únicamente el acopio de documentos encontrados sobre las universidades; el conocimiento que se generará de la investigación surgirá de un proceso sistemático de recolección de información que incluye simultáneamente la recogida y el análisis de esta, dando lugar a un ejercicio constante de generación de datos en donde tienen cabida ejercicios de tipo comparativo y muestreos no estadísticos.

4.2 Método de Investigación

Desde esta investigación se asume que la construcción de teoría sobre el objeto de estudio nace a partir de los mismos datos, esto es, del análisis documental que se haga sobre la información recogida y las diferentes etapas que se realicen para interpretarla. Por tal razón, se ha identificado que el método que favorece estas características es la teoría fundamentada. La TF es un método de investigación cualitativo en el que se busca construir teoría sobre un objeto de estudio a partir de los datos, antes que de la revisión de bibliografía especializada sobre ella. Para lograrlo se siguen procesos cíclicos de recolección de información y de comparación constante entre unidades de análisis, hasta el punto de llegar a una saturación teórica, esto es, que una nueva unidad de análisis no aporte nada nuevo a la conceptualización que emerge de los datos.

Siguiendo a Kolb (2012), este es un método desarrollado por Barney Glaser y Anselm Strauss, quienes creían que la teoría sobre un objeto de estudio podía emerger de la construcción de datos cualitativos, no tanto de la revisión de teoría especializada o datos previamente establecidos. Según Páramo (2015) la teoría

fundamentada es el resultado de la confluencia de corrientes interpretativistas: el pragmatismo de Peirce y el interaccionismo simbólico de Mead. De acuerdo con Sandín (2003) el pragmatismo considera la experiencia como punto de partida y llegada de todo conocimiento y el interaccionismo simbólico parte de la premisa que el investigador debe acercarse al objeto de interés desde el punto de vista de los sujetos que estudia. Lo anterior es importante en esta investigación, pues se busca capturar las percepciones de los programas sobre el lugar que ocupa el análisis de datos en la formación de profesionales, desde los diseños que reflejan cómo visualizan la formación de futuros profesionales.

En de la Espriella y Gómez (2020) se menciona que el desarrollo de la TF se debe a los sociólogos Barney Glaser y Anselm Strauss luego de haber publicado en 1965 *Awareness of Dying*, una investigación inductiva en moribundos de hospitales. Glaser, proveniente de la escuela de Columbia, de corte eminentemente cuantitativo, adoptó términos como el de muestreo, codificación y variable para brindarles un nuevo significado; mientras que la influencia de Strauss nace del pragmatismo y el interaccionismo simbólico propios de la escuela de Chicago, de donde el método adopta su orientación y significado. De aquí que Denzin y Lincoln (2018) caracterizan a la TF como la combinación de la influencia de estas escuelas, que en la mitad del siglo XX competían por dominar la escena de la sociología estadounidense.

Los desacuerdos entre Glaser y Strauss dieron lugar a una serie de reconstrucciones de la TF. Una de las más notables se destaca con Strauss y Corbin (1990), quizá el de mayor uso en la actualidad por su carácter sistemático. También se destacan la de Charmaz, o denominada diseño constructivista y el

diseño emergente de Glaser. Más allá de las variantes que pueda adoptar el método, el proceso de la TF sigue pasos que incluyen la recolección de información, codificación, escritura de memos, categorización y comparación constante, los cuales se pasarán a explicar para efectos de la investigación.

4.3 Diseño metodológico

Para aplicar la teoría fundamentada se siguen ciclos que incluyen la recolección de datos y codificación de información relevante con el fin de consolidar una categorización que permita construir la teoría sobre los datos, en donde se aplican estrategias como el muestreo teórico y la comparación constante entre unidades de análisis a fin de llegar a un punto de saturación teórica en el que la inclusión de nuevos datos no contribuya significativamente a la teoría.

De aquí en adelante, se explicará en términos de la investigación cómo se materializa cada etapa del método para conseguir los objetivos que se han trazado, basándonos en autores como Strauss y Corbin (1998), Kolb (2012) y Denzin y Lincoln (2018).

4.3.1 *Recolección de información*

Según Strauss y Corbin (1998) los datos pueden provenir de una gran variedad de fuentes como entrevistas, observaciones, documentos, grabaciones e incluso material que puede cuantificarse. Ante la imposibilidad de realizar entrevistas a profesores de Estadística en universidades, por el desinterés en general que se percibió con relación a una investigación preliminar acerca de la enseñanza de la Estadística -lo cual se mencionará en la sección de conclusiones-

, se acudirá a conseguir un repositorio documental con lo encontrado en los micrositios de los programas de pregrado.

Dentro de las fuentes de información que están incluidas se encuentran documentos oficiales sobre proyectos educativos del programa, brochures digitales, información que puede ser transcrita directamente del micrositio web del programa, videos grabados por profesores y/o estudiantes sobre el pregrado y entrevistas a egresados.

Se asume que la selección de la muestra no se hará pensando en conformar un grupo representativo de los programas de pregrado colombianos desde un punto de vista estadístico sino dirigido a la construcción de teoría. De cada pregrado se busca la mayor cantidad de información posible disponible en la web de lo que se busca en términos de los objetivos de la investigación -lo cual se explicará más adelante- y su tamaño dependerá de alcanzar un nivel de información que se consolide la teorización sin necesidad de añadir una unidad de análisis adicional.

4.3.2 Codificación

De acuerdo con Denzin y Lincoln (2018) la codificación es un proceso que permite asignar etiquetas a segmentos de texto para resumir, categorizar y en definitiva hacer posible la sistematización de la información. Estos autores asumen una visión constructivista de la teoría fundamentada -según Kathy Charmaz- y mencionan que de acuerdo con ella se hacen al menos dos tipos de codificación; la codificación inicial y la codificación enfocada.

La codificación inicial, que en algunos casos se denomina codificación abierta (Strauss y Corbin, 1998; Kolb, 2012) está orientada a identificar respuestas a preguntas que se hacen directamente de los datos e identificar lo que ellos dicen o dejan de decir. La codificación enfocada está más orientada al encuentro de una primera generación de categorías capaces de sintetizar en forma suficiente los códigos encontrados inicialmente y que serán objeto de un refinamiento efectuado a partir de un continuo ejercicio de comparación entre observaciones.

4.3.3 *Sobre la Primera Unidad de Análisis*

Ya que los sujetos de la población son los PECD -algo que se explicará más adelante en la sección 4.4- se entiende por primera unidad de análisis como el primer sujeto de este conjunto. A continuación, se explica cómo se lleva a cabo su análisis.

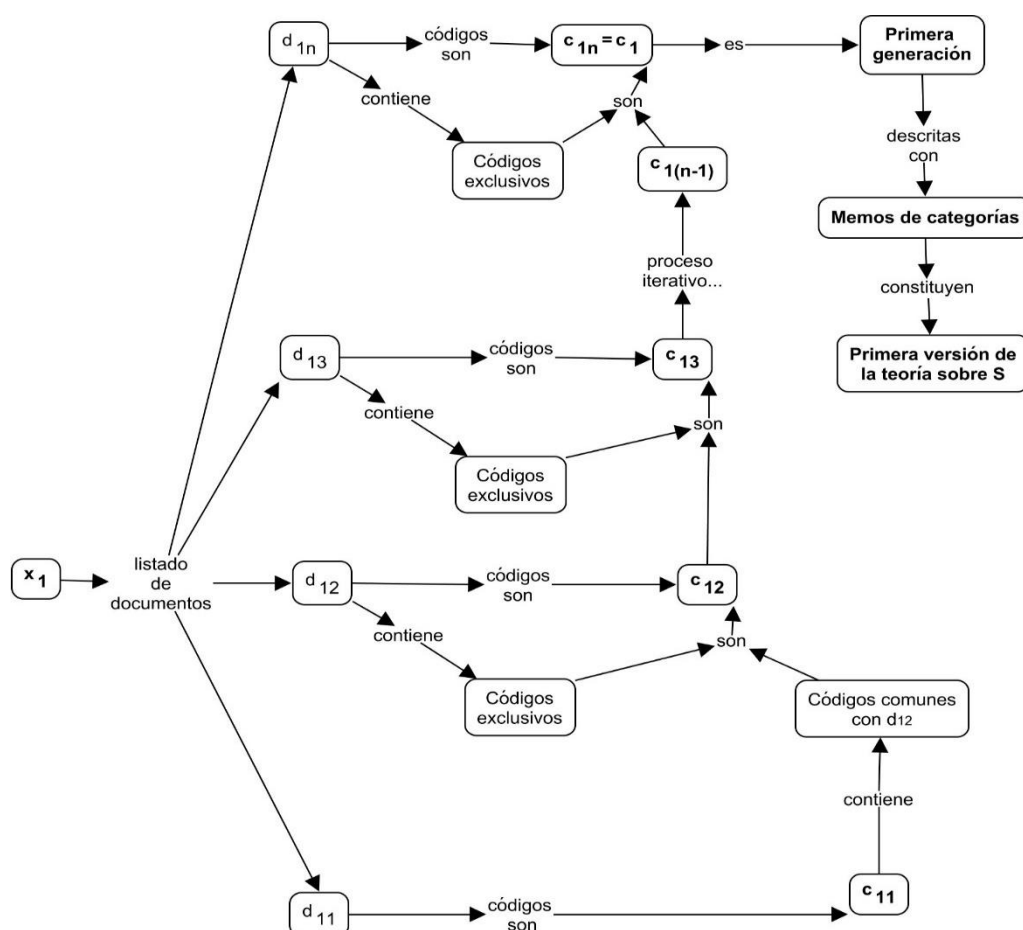
Para garantizar coherencia metodológica y buscando que la teoría emerja de la información recolectada, en esta investigación se realizará codificación inductiva, proceso que parte de la información recogida sobre cada unidad de análisis y que tiene por objeto su categorización a partir de la identificación de conceptos clave que surgen de la revisión documental, no de una lista predeterminada de ideas del investigador o de sus preconceitos sobre el tema que busca investigar.

Para llevar a cabo la codificación inductiva se adopta la visión constructivista de la TF según Strauss y Corbin (1998), para los cuales se deben hacer al menos dos tipos de codificación; la codificación inicial y la codificación enfocada. La codificación inicial -que en algunos casos se denomina codificación

abierta- está orientada a identificar respuestas a preguntas que se hacen directamente de los datos e identificar lo que ellos dicen o dejan de decir. La codificación enfocada está más orientada al encuentro de una primera generación de categorías capaces de sintetizar en forma suficiente los códigos encontrados inicialmente y que serán objeto de un refinamiento efectuado a partir de un continuo ejercicio de comparación entre observaciones. La Figura 1 pretende mostrar lo que se hace con x_1 , la primera del conjunto de unidades de análisis $S = \{X_1, X_2, \dots, X_n\}$.

Figura 1

Tratamiento de la primera unidad de análisis



La codificación inductiva comienza con la recolección de toda evidencia disponible en los micrositos web a través de su transcripción en ATLAS.Ti 22 para consolidar una lista de n documentos asociados con la primera unidad de análisis. El ciclo de TF asociado con x_1 consiste en un proceso iterativo que se hace con cada uno de sus n documentos $\{d_{11}, d_{12}, \dots, d_{1n}\}$ asociados que incluye los procesos de codificación abierta, enfocada y construcción de teoría.

Con el primer documento d_{11} el proceso de codificación abierta consiste en establecer citas libres con sus ideas principales y la identificación de códigos que están contenidos en cada una de ellas. Con la lista de códigos que se genera se realiza la codificación enfocada, la cual consiste en encontrar relaciones entre los que resulten más recurrentes y con ello describirlos en términos de ellas. Entonces, surge la primera lista de códigos clasificable por criterios como la repitencia o la cantidad de relaciones encontradas, esto es, la densidad. El proceso inductivo asociado con d_{11} finaliza con el listado c_{11} de códigos más densos de x_1 y su descripción a través de memos que describen a sus relaciones.

Con el fin de enriquecer las primeras versiones de los memos anteriormente mencionados, el ciclo asociado con x_1 continúa con el análisis de d_{12} , siguiendo procesos análogos de codificación abierta y enfocada. Así mismo, se encuentra el segundo listado de conceptos clave que puede ratificar algunos de los códigos más relacionales del documento 1 o ampliarla con aquellos que surjan de la revisión exclusiva de documento 2. Esta comparación permite enriquecer la descripción comenzada con los memos por categoría, el memo de x_1 , y consolidar un listado de códigos más relacionales asociados con los dos primeros documentos.

Finalmente, se hacen tantas iteraciones del proceso anterior como documentos asociados hayan de x_1 haciendo comparación constante entre el listado $c_{1(n-1)}$ de códigos más relacionales de sus primeros $n-1$ documentos con los códigos nuevos de d_{1n} -el n -ésimo documento-. La primera generación de categorías, de aquí en adelante c_1 , es el conjunto de códigos más relacionales al final del análisis de los n documentos asociados con x_1 . El primer ciclo finaliza con la consolidación de los memos por categoría y memo de x_1 , que es resultado de la comparación constante entre los N documentos mencionados. Los memos anteriormente mencionados constituyen la primera versión de la teoría sobre S .

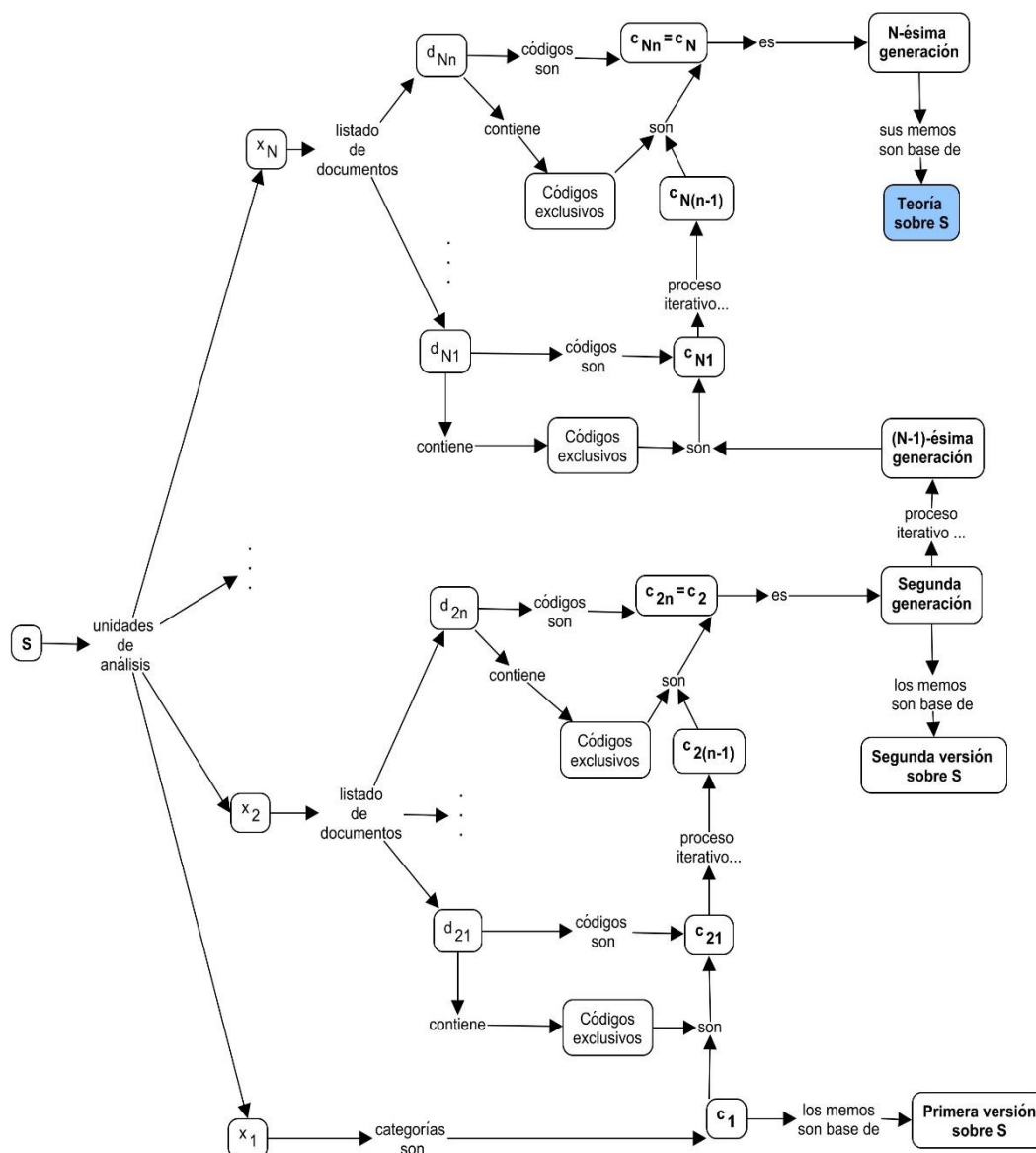
4.3.4 Sobre las Unidades de Análisis Posteriores

Para continuar con el proceso de teorización sobre el conjunto $S = \{x_1, \dots, x_N\}$ de unidades de análisis de la población de estudio, el ciclo de TF explicado en la sección anterior se repite $N-1$ veces. La Figura 2 muestra los demás ciclos del proceso de TF.

En el caso de x_2 -la segunda unidad de análisis- se toma como base a c_1 y la primera versión de teoría sobre S . Los procesos de codificación abierta y enfocada conducentes a refinar a c_1 se llevan a cabo con d_{21} -el primer documento de x_2 - y se replican para los documentos $\{d_{22}, \dots, d_{2n}\}$ para obtener como producto la segunda generación de categorías, que reflejan el proceso de comparación constante y permiten una descripción más elaborada que lleva a la segunda versión de la teoría sobre S .

Figura 2

Ciclos iterativos de TF para las demás unidades de S



Las revisiones iteradas con las demás unidades de análisis se agotan hasta llegar a un punto de saturación teórica, en el cual no es necesario analizar más documentos. El ciclo final de TF se hace con x_N -la última unidad de análisis- en la que se parte de la (N-1)-ésima versión de la teoría sobre S a partir de su correspondiente generación de categorías; se analiza el documento d_{Nn} para

obtener las categorías más importantes de S, junto con su descripción, la que constituye la teoría sobre S, el conjunto de unidades de análisis.

4.3.5 Triangulación Intra-método y Triangulación de Datos

Con el fin de alcanzar validez de la información que se recoja de cada programa de pregrado, se han aplicado diversas técnicas de recolección de datos: en un primer momento, el acopio de documentos oficiales de los pregrados o transcripciones de información en el micrositio web del programa; a continuación, el análisis de videos de diversidad de actores en el análisis de datos; y finalmente, la elaboración de entrevistas a egresados de los programas. En todos los casos, la TF fue usada como el método para construir teoría, con lo que la obtención de memos y categorías siguió un proceso de triangulación intra-método, el cual, según Alzás y Casa García (2017) “consiste en analizar los datos utilizando un solo método, pero seleccionando diversas técnicas de recogida de información enmarcadas dentro de la línea estratégica de dicho método.” (p. 405).

No obstante, para caracterizar al científico de datos se transcribió información directamente de páginas web, pero se cuenta con videos que explican lo que hace profesionalmente, o generalidades de los pregrados; dentro de los que se incluye la visión de un profesional en Ciencia de Datos, un profesor y un estudiante, lo que permite tener diversas fuentes de información y visiones distintas del mismo fenómeno. Por tal razón, el análisis de resultados relativos al científico de datos tuvo lugar con base en una triangulación de datos en el espacio, que en términos de Alzás y Casa García (2017) implica la validez de una

proposición con relación a un fenómeno desde los aportes de una población diversa.

Entre tanto, para caracterizar al estadístico, aunque se usaron fuentes de información parecidas al caso del científico de datos, la línea de tiempo que se tuvo en cuenta es mucho más amplia; pues se cuenta con propuestas con año de creación de 1958 hasta 2020, con lo que se puede validar o contrastar afirmaciones hechas sobre un fenómeno en momentos temporales distintos (Alzás y Casa García, 2017). En ese sentido, el análisis de resultados se hizo con base en una triangulación de datos en el tiempo.

4.4 Sujetos de la Investigación e Información Disponible en la Web

En términos de los objetivos de la investigación se buscan analizar los programas nacionales de pregrado en Estadística o Ciencia de Datos (PECD) activos a diciembre de 2021 según SNIES, a la luz de lo que hacen para garantizar la formación en la competencia de análisis de datos, para lo cual se ha pensado en dividirlos en tres grupos que son (a) programas de pregrado en Estadística (PE), (b) programas de pregrado en ciencia de los datos (CD) y (c) programas de pregrado en ingeniería en ciencia de los datos (ICD).

Para el momento de la consulta, la base de datos del SNIES reportó 14 programas de pregrado en Estadística (PE), de los cuales se escogieron 8 programas de pregrado en Estadística a nivel nacional en las siguientes universidades:

- Universidad Nacional de Colombia Sede Bogotá (UN)

- Universidad del Valle (UV).
- Universidad Santo Tomás (USTA).
- Universidad de Antioquia (UA).
- Universidad INCCA de Colombia (INCCA).
- Fundación Universitaria los Libertadores (FUL).
- Universidad ECCI (ECCI).
- Universidad el Bosque (UB).

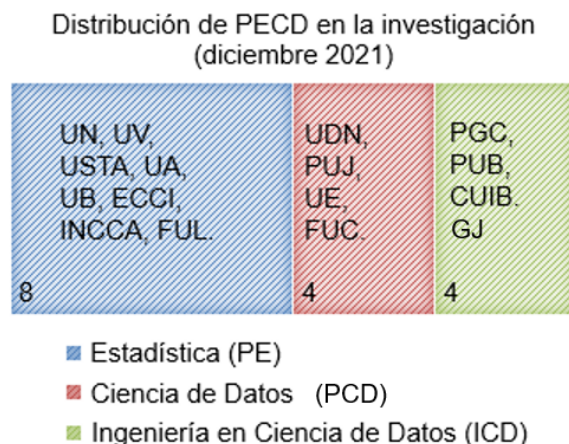
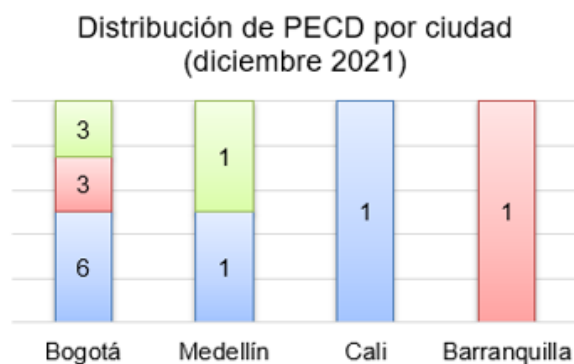
La UN oferta tres pregrados en Estadística, todos con ofertas similares, por lo que se tomó la decisión de incluir en la investigación solamente al de Bogotá y descartar los de la Sede Medellín y la Sede La Paz. Así mismo, por tratarse de denominaciones híbridas se descartaron del análisis final los programas de Estadística y Ciencias Actuariales de la Fundación Universidad de América, y el programa de Matemáticas con énfasis en Estadística de la Universidad del Tolima. El programa de Ingeniería Estadística de la Escuela Colombiana de Ingeniería Julio Garavito (JG) se incluyó dentro del análisis llevado a cabo con los programas de Ingeniería en Ciencia de Datos. Para el momento de la recolección de información, el programa de Estadística de la Universidad de Córdoba no tenía ninguna información en su sitio web oficial.

Entre tanto, SNIES reportó 9 pregrados de Ciencia de Datos (CD): cinco se ofrecen como programas de pregrado en Ciencia de Datos, tres como programas de pregrado en Ingeniería en Ciencias de Datos y el restante es el programa de

pregrado de Matemáticas Aplicadas a la Ciencia de Datos –ofertado por la Corporación Universitaria Autónoma del Cauca (CUAC)–. Se incluyó en la investigación a cuatro programas de pregrado en Ciencia de Datos: Universidad del Norte (UDN), Pontificia Universidad Javeriana (PUJ), Universidad Externado de Colombia (UE) y Fundación Universitaria Compensar (FUC), que ofrece su programa en dos modalidades -presencial y virtual-. Para la investigación se incluyeron las cuatro universidades oferentes, pero los dos programas de la FUC se trataron como uno solo, por tratarse de ofertas que solo difieren en su modalidad pero que sus contenidos son idénticos.

También se analizaron los tres programas de ingeniería en Ciencia de Datos (ICD): Politécnico Gran Colombiano (PGC), Pontificia Universidad Bolivariana (PUB) y Corporación Universitaria Iberoamericana (CUIB). No se incluyó el programa de CUAC porque al momento de la consulta no contaba con información sobre el programa de su pregrado. Para efectos de la investigación, el programa de Ingeniería Estadística de la Escuela Colombiana Julio Garavito se incluyó dentro del grupo de programas de Ingeniería en Ciencia de Datos, ya que su oferta académica es más compatible con un programa en Ingeniería en Ciencia de Datos que de Estadística; llegando a un total de cuatro unidades de análisis disponibles.

La Figura 3 muestra un resumen de la información anterior, mientras que la Figura 4 da cuenta de la ubicación geográfica de los PECD:

Figura 3*Distribución de los PECD***Figura 4***Distribución de PECD por ciudad de origen*

Se puede ver que el 75% de los PECD elegidos son ofertados en Bogotá; también, 75% de los PE, 75% de los CD y 75% de los ICD son de Bogotá. En Medellín se ofertan dos PECD (12.5%); el de Estadística en la UA y el de Ingeniería en Ciencia de Datos de la PUB. En Cali se oferta el pregrado en Estadística de la UV, y el Barranquilla el de Ciencia de Datos de la UDN.

La Tabla 1 da cuenta de otros elementos que caracterizan a los PE elegidos; entre ellas, la facultad que los oferta, el año de creación, la duración del programa (t , en semestres) y el tipo de reconocimiento que el ministerio le ha otorgado. En el caso de INCCA la duración se da en cuatrimestres.

Tabla 1

Información general de los PE

Universidad	Facultad oferente	Año de creación o funcionamiento	t	Reconocimiento del ministerio
UN	Facultad de Ciencias	1958	9	Acreditación de alta calidad
UV	Facultad de Ingeniería	1979	10	Acreditación de alta calidad
USTA	Facultad de Estadística*	2005	8	Registro Calificado
UA	Facultad de Ciencias Exactas y Naturales	2011	9	Registro Calificado
UB	Ingenierías y Administración	2014	9	Registro Calificado
ECCI	Facultad de Ciencias Económicas y Administrativas	2014	9	Registro Calificado
INCCA	Departamento de Matemáticas, Física y Estadística	2019	9	Registro Calificado
FUL	Facultad de Ingeniería y Ciencias Básicas	2020	8	Registro Calificado

Como se puede apreciar, los dos programas que llevan más tiempo ofertando son los únicos con Acreditación de Alta Calidad (UN, UV). Tres de los PE los ofertan facultades de ciencias o afines (UN, UA, INCCA), dos por

facultades de ciencias económicas o afines (USTA, ECCI), el de UV es el único ofertado por una facultad de Ingeniería. También, dos son ofertados por facultades híbridas o que no pertenecen a una sola área de conocimiento: FUL, ofertado por la facultad de Ingeniería y Ciencias Básicas; y UB, ofertado por la facultad de Ingenierías y Administración. La mayoría de los PE cuentan con una estructura curricular diseñada para 9 semestres (4 de 8: 50%); dos, para 8 semestres; uno, para 10 semestres y el de INCCA para 3 años. Cabe señalar que INCCA y FUL no cuentan con estudiantes graduados hasta el momento.

A continuación, la Tabla 2 muestra información de los PCD análoga a la de la tabla anterior:

Tabla 2

Información general de los PCD

Universidad	Facultad oferente	Año de creación o funcionamiento	T	Reconocimiento del ministerio
PUJ	Ciencias, Comunicación y Lenguaje, Ingeniería, Ciencias Económicas y Administrativas	2020	8	Registro Calificado
FUC	Facultad de Ingeniería	2020	9	Registro Calificado
UDN	División de Ciencias Básicas	2021	8	Registro Calificado
UE	Departamento de Matemáticas	2022	8	Registro Calificado

Como se puede apreciar, ninguno de los CD tiene estudiantes egresados y todos cuentan con registro calificado vigente. Dos de los CD los ofertan facultades de ciencias básicas o afines y el de FUC lo oferta la Facultad de Ingeniería. En el caso de PUJ, el programa fue diseñado por las cuatro facultades mencionadas, las cuales funcionan en forma independiente en la Universidad.

De manera análoga, la Tabla 3 muestra información sobre los ICD de la investigación; habiendo aclarado que el programa de GJ hará parte de este grupo. El programa de CUIB es en modalidad virtual.

Tabla 3

Información general de los ICD

Universidad	Facultad oferente	Año de creación o funcionamiento	t	Reconocimiento del ministerio
PUB	Facultad de Ingeniería	2019	9	Registro Calificado
GJ	Departamento de Matemáticas, programa de Ingeniería de Sistemas	2021	10	Registro Calificado
PGC	Facultad de Ingeniería, Diseño e Innovación	2021	9	Registro Calificado
CUIB*	No disponible	No disponible	8	Registro Calificado

Hay menos información disponible en sitios web. No se dispone de información en CUIB con relación a la facultad que oferta el pregrado ni su año de creación o funcionamiento. El caso de GJ es el resultado del trabajo conjunto entre el pregrado en Matemáticas y el de Ingeniería de Sistemas, parecido a lo que ocurrió con el de Ciencia de Datos en PUJ. Con relación al de PGC, es el resultado de una línea de trabajo como la observada en los pregrados en Estadística de FUL y UB; facultades con denominaciones híbridas ofertando el programa.

Para terminar esta sección, se procederá a mostrar la información conseguida en los micrositios web de cada PECD en la Tabla 4. Se dispone de una variedad más alta de información en el caso de los PE, aparte de contar con los PEP de algunos de ellos, también se resaltan elementos misionales y visionales, y una presentación de programa. La tendencia a usar videos para promocionarse como programas es más frecuente en PCD o ICD; mientras que en caso de los PE solo una tiene información en esta clase de formato –ECCL, de estudiantes de la carrera que se visualizan a futuro–.

Probablemente, por lo reciente de las propuestas de programas en Ciencia de Datos o afines, la información extraída de los PCD o ICD sea mucho menor que en la de los PE. Los espacios en blanco indican que la información señalada no estaba disponible.

Tabla 4*Información disponible en la web por cada PECD*

Univ.	Perfiles	¿Por qué estudiar la carrera?	Presentación de programa	Misión	Visión	PEP
UN	Aspirante y egresado	No	Sí	Sí	Sí	Sí
UV	Egresado	No	Sí	Sí	Sí	Sí
USTA	Estudiante, profesional y ocupacional	Sí	Sí	Sí	Sí	No
UA	Egresado	No	Sí	Sí	Sí	No
UB	Estudiante y egresado	Sí	Sí	No	No	No
ECCI	Egresado y ocupacional	No	Sí	No	No	No
INCCA	Aspirante, egresado y ocupacional	Sí	No	No	No	No
FUL	Proyección profesional	Sí	Sí	No	No	No
UDN	Egresado	No	Sí *	No	No	No
PUJ	Perfil de desempeño	No	Sí *	No	No	No
UE	Estudiante, egresado y ocupacional	Sí	Sí *	No		
FUC	Campo laboral	No	No	No		No
GJ	Aspirante y profesional	Sí	No	No		No

Univ.	Perfiles	¿Por qué estudiar la carrera?	Presentación de programa	Misión	Visión	PEP
CUIB*	Aspirante y profesional	Sí	No	No		No
PGC	Ocupacional	No	Sí	No		No
PUB	Áreas de desempeño	No	Sí	No		No

5. Resultados

"Científico de datos (n): Persona que sabe más de Estadística que cualquier programador y que a la vez sabe más de programación que cualquier estadístico".

(Wills, 2012)

En este capítulo se mostrarán los resultados de la aplicación de la TF en los sujetos de investigación. En concordancia con los objetivos específicos de la investigación, se muestra como los PECD manifiestan rasgos característicos de la Sociedad del Dato en sus ofertas; a continuación, se presentan caracterizaciones de cómo se visualizan los diversos tipos de profesionales que trabajan con datos, esto es, el estadístico, el científico de datos y el ingeniero en Ciencia de Datos. Finalmente, se pasará a identificar el lugar que los PECD le otorgan al análisis de datos.

5.1 Manifestaciones de la Sociedad del Dato según PECD

Para dar cumplimiento al primer objetivo específico de la investigación, se ofrece a continuación una descripción detallada de cuáles elementos descritos en la caracterización de la Sociedad del Dato (ver capítulo 3) se reflejan en cada uno de los PECD. En ese sentido, se dará a conocer lo identificado en los programas de pregrado en Estadística, luego los de Ciencia de Datos y, por último, los de Ingeniería en Ciencia de Datos. Se ha encontrado que, en mayor o menor medida, los PECD admiten un escenario de producción masiva de datos, marcado por el

fenómeno del Big-Data; y conceptualizaciones sobre la Estadística y la Ciencia de Datos, campos de conocimiento que nutren a una sociedad vista desde un lente data-céntrico.

5.1.1 *Programas de Estadística*

Desde los PE se pueden ver en alguna medida aspectos relacionados con el contexto de reconocimiento de la Sociedad del Dato, donde salen a relucir el hecho de la producción masiva de datos y la responsabilidad ética en el trabajo con datos. Así mismo, por ser la Estadística una disciplina que nutre a la Sociedad del Dato, cada programa de pregrado ofrece una visión a partir de intentos por conceptualizar, y reconocer las áreas donde esta se encuentra presente.

5.1.1.1 *La Sociedad del Dato según PE.*

Desde algunos PE se reconoce la producción masiva de datos como una realidad que permea a la sociedad en general y, por tanto, la labor del estadístico; no obstante, solamente la UV acude al término Big-Data para hacer referencia a este fenómeno. La UV llama la atención sobre la necesidad de tratar bases de datos textuales, que escapan al dominio de las herramientas clásicas de la Estadística. ECCI, FUL y UV comparten que el profesional en Estadística está expuesto continuamente a situaciones donde debe enfrentarse a grandes volúmenes de datos.

Por otra parte, existen diversas maneras de instalar un contexto reinante en la sociedad en general. Mientras la UB visualiza un contexto que parte del siglo XX, posicionando a la Estadística como la FUL presenta su programa mostrando

un contexto como el de la Revolución 4.0, o, en otras palabras, del acelerado crecimiento tecnológico, el aumento en la velocidad de generación de información y la cantidad de datos disponibles.

Un aspecto que es preocupación en medio de la Sociedad del Dato es la dimensión ética del uso que se le brinda a los datos. Precisamente, la ética es una categoría relevante en el análisis hecho con TF, tal y como se aprecia en la Figura 5, donde se aprecia el número de veces que aparece esta categoría en cada uno de los PE.

Figura 5

Distribución de la categoría ética en los PE



Aunque algunos programas solo mencionan la ética como un elemento que es parte de la formación de sus profesionales, otros ponen acento en que es un elemento intrínseco en el tratamiento con datos. En La UV menciona que la ética es un rasgo de la calidad con la que se espera que se haga el ejercicio profesional y que soporte todo uso de bancos de datos. De USTA se puede inferir que, en el proceso de generación de conocimiento y difusión de este, en el cual los datos

median, existe una fuerte carga ética. La UN afirmó la necesidad de respeto de la ética Estadística frente a los datos.

5.1.1.2 La Estadística según PE.

Con relación a la búsqueda de elementos que ayuden a capturar la concepción de los PE sobre la Estadística como campo de conocimiento, solo la mitad de ellos (UB, UV, USTA, FUL) dan ideas acerca de su objeto de estudio y contexto de desarrollo. Ninguno de ellos declara abiertamente que la Estadística sea una ciencia, sino que le confieren un estatus como disciplina (UB), herramienta en el método científico (UV) o como una actividad de base científica (USTA).

De acuerdo con el programa de la UB, se declara que la Estadística es una disciplina del siglo XX orientada a procesos que giran en torno a los datos: recolección, análisis e interpretación, así como al estudio de fenómenos aleatorios y la toma de decisiones. La UV vincula la Estadística al hallazgo de resultados reproducibles y confiables. La USTA declara que su objetivo es el tratamiento de información tanto cualitativa como cualitativa. Entre tanto, según la FUL (Fundación Universitaria los Libertadores) la Estadística es una ciencia del futuro gracias al contexto de la revolución 4.0, caracterizada por un acelerado crecimiento tecnológico y una rápida generación tanto de datos como información.

Con respecto a los campos de conocimiento en los que la Estadística tiene aplicación, enunciados en los PE, parece haber un acuerdo en que las circunstancias políticas, sociales y económicas han impulsado su uso; a tal punto que las áreas de influencia son demasiadas como para ser mencionadas

totalmente. Sin embargo, y sin llegar a mencionarlas en orden de importancia, varios de los programas nombran explícitamente algunas, tal y como se resume en la Tabla 5.

Tabla 5

Áreas de aplicación de la Estadística reconocidas en los PE

Universidad	Áreas de influencia
UN	Academia, salud, industria y servicios, finanzas, investigación de mercados, sector público, investigación agropecuaria.
UV	Administración, biología, ciencias sociales, demografía, docencia universitaria, economía, educación, investigación básica en Estadística, investigación de mercados, medicina, negocios, psicología, salud, sector bancario y seguros.
USTA	Política, economía, salud, psicología, ingeniería, finanzas, mercadeo, biología, medicina y sociología.
UA	Administración pública, industria, servicios, docencia, ciencias de la salud, ciencias naturales, ingeniería, economía y finanzas.
UB	Medicina, economía, psicología, geología, astronomía, física.

De acuerdo con lo anterior, las áreas de aplicación más mencionadas por los documentos son la salud, la economía y la investigación. Con relación a la salud, la UV menciona escenarios como el “análisis de costos y frecuencias de accidentalidad, enfermedad e invalidez, problemas con la atención médica, la hospitalización y el seguro de la salud pública” (Universidad del Valle, 2017, p. 8). Sobre la economía, “la medición de la producción de bienes y servicios, precios, volúmenes de las transacciones, todo tipo de recursos, fuerza laboral y niveles de vida. Reacciones de los mercados frente a cambios de precio, publicidad y

legislaciones gubernamentales” (Universidad del Valle, 2017, p. 8). Respecto de la investigación, abarca no solamente la que se hace en Estadística sino la efectuada en mercadeo, el sector agropecuario y las ciencias en general.

Así mismo, la UV reconoce que diferentes organismos nacionales e internacionales usan técnicas y herramientas Estadísticas para sus investigaciones y ejercicios de toma de decisiones, colocando ejemplos de organismos como CEPAL, OCDE, BID, DANE y Fedesarrollo. También, que, a nivel productivo y empresarial, los bancos, ingenios, instituciones de educación superior, centros de investigación y entidades gubernamentales en general, requieren de los conocimientos del profesional en Estadística para hacer más eficientes sus procesos de producción de información.

Mención especial merece que solo dos de los PE reconocen que la docencia es una de las áreas donde se aplica la Estadística: en la UV mencionan específicamente su aplicación en la enseñanza de técnicas Estadísticas y la investigación en universidades, mientras que en UA lo mencionan sin brindar mayores detalles. Esto ratifica resultados de investigaciones con relación al profesor de Estadística, donde se concluye el uso limitado que hacen de ella (Watson, 2001; Makar y Confrey, 2005; Stohl, 2005) y su poca familiaridad con el carácter no determinista del análisis de datos (Estrella, 2017), algo que a menudo interfiere con los procesos de enseñanza-aprendizaje de conceptos estadísticos a nivel universitario (Rocha, 2012).

5.1.2 *Programas en Ciencia de Datos*

De acuerdo con los PCD y los elementos que resaltan de la Sociedad del Dato en ellos, las propuestas parecen estar acorde con un mundo donde la producción de datos es cada vez más masiva, con implicaciones en la forma como el hombre y la máquina se relacionan y nuevos retos para la sociedad en general. Así mismo, mencionan ejemplos cercanos a la realidad de las personas para ilustrar el impacto que la Ciencia de Datos tiene en ellas, lo que refleja comportamientos propios de una sociedad data-céntrica, caso puntual la producción de datos en redes sociales. También, se llama la atención por hacer del trabajo con datos una labor más ética, en el sentido de procurar el bien general de la sociedad.

Finalmente, al considerar que la Ciencia de Datos es un campo disciplinar que ayuda a moldear la Sociedad del Dato, se brindan descripciones de este campo de conocimiento.

5.1.2.1 *La Sociedad del Dato según PCD.*

Los PCD sugieren desde sus descripciones a una sociedad permeada por los datos. Por ejemplo, la UDN menciona de ellos que

[...] son la base de todo. Cada segundo se generan miles de millones de datos de diferente naturaleza en el mundo. Hemos acumulado una gran cantidad de información en forma de datos y seguimos produciéndola. El reto al que nos enfrentamos como sociedad es poder entender y usar ese universo de información para la generación de nuevos conocimientos, la

gestión de proyectos y tomar de decisiones que transformen nuestra sociedad. (parr. 1).

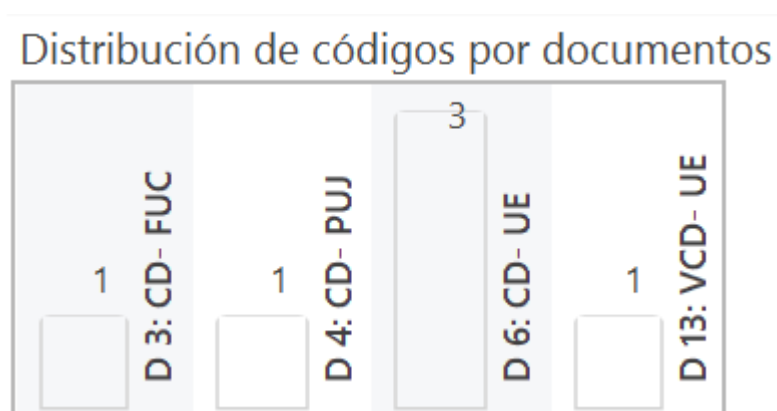
El suceso de la masiva producción de datos es también reconocido por los demás PCD. La UE en su video promocional señala que en los últimos años se ha producido la humanidad ha producido la mayoría de los datos con los que cuenta. FUC y PUJ sitúan este fenómeno en un contexto más amplio como el de la cuarta revolución industrial; FUC profundiza más en esto en su documento maestro, al advertir que una de sus secuelas es la de un mundo interconectado, donde la relación hombre-máquina se ha vuelto más estrecha con el tiempo, gracias a la incursión del dato como elemento potenciador.

Adicionalmente, algunos programas acuden a situaciones de la vida real, propias de la Sociedad del Dato, que son atribuibles al influjo de la ciencia de los datos. La UE menciona en su video promocional caso como la predicción que hacen los algoritmos presentes en las redes sociales para detectar preferencias de los consumidores y posibles círculos sociales; y la necesidad de compartir información que ahora se ha vuelto dato (fotos, videos, libros on-line que se consumen, entre otros). Así mismo, en su información web añade la modelación de riesgo financiero, la optimización de recursos públicos y la predicción de factores de consumo. PUJ en su video promocional recuerda que áreas de desempeño como la salud –especialmente a raíz de los datos generados por el desarrollo del COVID-19–, el derecho –especialmente en el análisis de crímenes y de tutelas–, la biología, la economía, la banca, y el periodismo.

Por otra parte, existe una preocupación por la calidad y la ética con la que se manejan los datos y la información conseguida de ellos. La Figura 6 muestra el número de veces que aparece en cada programa analizado y que esta es importante en al menos tres de los PCD.

Figura 6

Distribución de la categoría ética en los PCD



PUJ destaca que dentro del proceso de formación debe imprimirse una capacidad para reflejar una visión ética del ejercicio profesional. En UE, la dimensión ética proviene del reconocimiento del poder de los datos en la toma de decisiones y el impacto en la humanidad. FUC añade al aspecto ético, procurar proteger la información que se use.

5.1.2.2 La Ciencia de Datos según PCD.

Aunque el análisis efectuado reporta que la Ciencia de Datos es un concepto ampliamente mencionado en los documentos de los programas analizados, no hay una definición explícita de lo que asumen como definición. Lo más cercano a una conceptualización se consigue de PUJ, universidad que ofrece

una propuesta multidisciplinar por la confluencia de cuatro facultades ofertantes y de donde puede inferirse que es un campo interdisciplinar para la solución de problemas representados por datos, que permite entender la influencia que tienen en la vida de las personas y sus implicaciones éticas. La UDN, hace referencia al Ministerio de las Tecnologías de la Comunicación para afirmar que la Ciencia de Datos es una de las habilidades más importantes que se deben lograr en la población colombiana.

Existen variedad de percepciones acerca de las áreas de conocimiento que dan forma a la CD. De la información de la UDN puede extraerse que hay cuatro componentes que le dan forma: (a) la Estadística, como base de lo que se hacía en los comienzos de la CD o la teoría de la decisión; (b) Machine Learning, o aprendizaje de máquina guiado a través de algoritmos hechos por el hombre; (c) inteligencia artificial, o la automatización de un número considerable de reglas en corto tiempo y (d) Deep Learning o redes neuronales, algoritmos diseñados para procesamiento a grandes escalas. La UE sugiere que la CD integra la matemática, la Estadística y la computación para convertir los datos en conocimiento y resolver problemas reales.

Entre tanto, la FUC en su documento maestro caracteriza a la Ciencia de Datos bajo tres pilares: la omnipresencia de los datos; el desarrollo de tecnologías para hacer frente a los desafíos que implican el Big-Data; y las personas, como los actores que definen el destino de todo intento por modelar, validar y analizar lo que se haga con los datos. Así mismo, menciona que en Ciencia de Datos se usan fundamentos de Machine Learning, Inteligencia Artificial y Analítica de Datos para la detección de patrones, la optimización de procesos y la elaboración de

predicciones; mediante herramientas como las Tecnologías basadas en Sistemas de Información y en la nube, acciones que permiten entender y analizar diferentes escenarios en los que los datos tienen protagonismo.

5.1.3 *Programas de Ingeniería en Ciencia de Datos*

Al igual que en los otros dos tipos de programas –especialmente los PCD–, desde lo ofertado en los ICD pueden visualizarse elementos propios de la Sociedad del Dato como la producción masiva y desbordante de datos. También, del mismo modo que en los PCD, se presenta a la Ciencia de Datos como la disciplina regente, con lugar propio en el dominio de otras disciplinas y que tiene sus propios ámbitos de aplicación.

5.1.3.1 *La Sociedad del Dato según ICD.*

Los ICD reconocen escenarios marcados por la necesidad de recolectar y aprovechar el potencial que tienen los datos. En el caso de GJ, el marco es la revolución 5.0. caracterizada por la irrupción de la Ciencia de Datos como herramienta para la búsqueda de conocimiento mediante la explotación de datos, y la transformación digital. CUIB menciona al Big-Data, haciendo énfasis más que todo en la producción de datos a gran escala; algo que, de lo visto en el video promocional en PUB, puede inferirse que es una labor emprendida a finales de siglo pasado, tanto en sectores públicos y privados.

El elemento ético del manejo de datos, muy presente en programas de PE y PCD, no se menciona de la misma manera en los ICD; únicamente la CUIB afirma

que debe estar presente en los ejercicios de toma de decisiones y en las propuestas de soluciones integrales a los problemas de la sociedad en general.

5.1.3.2 ¿Ciencia de Datos o Ingeniería en Ciencia de Datos?

Aunque el conjunto de las ICD tiene una denominación que cabría dentro del grupo de programas de pregrado en Ingeniería, la disciplina de base no es netamente ingenieril. Quizá por tratarse de una denominación distinta, el programa en GJ menciona como disciplina de base a la ingeniería Estadística, presentándola como una fusión entre Ciencia de Datos y transformación de tecnologías en función del análisis de información; pero en los demás casos, no se observa algo distinto a intentos por describir lo que entienden por Ciencia de Datos.

En ese orden de ideas, cada universidad muestra una idea distinta con relación a lo que entienden por Ciencia de Datos. PGC define a la Ciencia de Datos como “una disciplina que busca analizar de manera descriptiva, predictiva y prescriptiva las fuentes de información consideradas datos y que pueden ser bases numéricas, análisis de texto, imágenes y contenido audiovisual”. (parr. 1).

PGC, PUB y GJ coinciden en afirmar que la Estadística y elementos relacionados con la computación o la programación son herramientas que dan forma a la Ciencia de Datos; pero cada una añade a esta lista elementos distintos. Además de las anteriores, PGC nombra las matemáticas, inteligencia artificial, machine learning, ingeniería de datos y Big-Data; PUB, la programación de software, inteligencia de negocios y analítica de datos; GJ reconoce el valor de las matemáticas aplicadas y la computación orientada al análisis científico de datos.

Al igual que en los PCD, los ICD identifican algunos ejemplos concretos de aspectos que pueden ser abordados desde la Ciencia de Datos. Por ejemplo, en el video promocional de la PUB destacan el análisis de sentimientos y comportamientos en redes sociales, la automatización a través del internet de las cosas, visión artificial y modelos predictivos a partir de Machine Learning y elaboración de informes mediante inteligencia de negocios, entre otras. PGC habla de áreas de conocimiento como mercadeo, tecnología, deportes, finanzas y economía. GJ es aún más general en su mención: la ciencia, la academia y las empresas.

5.2 El Profesional del Trabajo con Datos según PECD

Cada una de las universidades oferentes de PECD modela una clase de profesional experto en manejo de datos que se resume en los diversos perfiles que se dan a conocer (ocupacional, egresado, profesional), las competencias y/o habilidades que se espera desarrollar a lo largo de la carrera, y los valores que se esperan formar. Estas características están permeadas por elementos como las facultades que ofertan estos programas, el tiempo que llevan funcionando, la localización geográfica de los mismos, entre otros aspectos, que se convierten en aspectos que explican las diferencias encontradas entre lo declarado en los PECD.

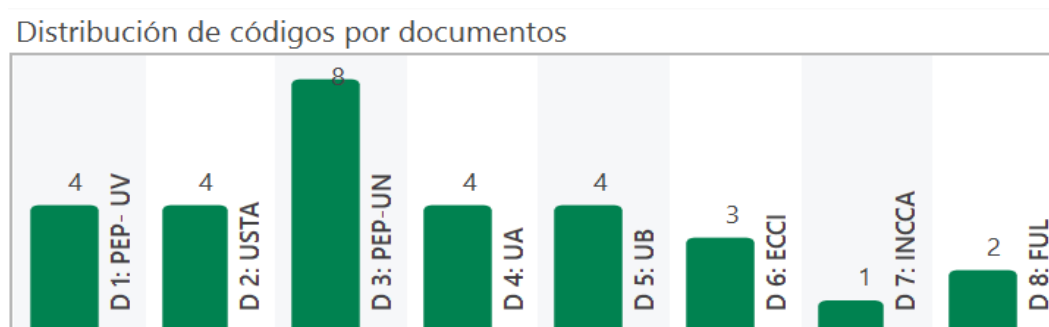
Con el ánimo de dar cumplimiento al segundo objetivo, y guardando una estructura similar a de la sección 5.1., se hará una descripción de lo encontrado en los PECD sobre los estadísticos, los científicos de datos y los ingenieros en Ciencia de Datos.

5.2.1 El Estadístico como Profesional del Dato

Dentro del análisis que se llevó a cabo con la TF se observó que la categoría *Estadístico* es la más densa entre los documentos de los PE. La Figura 7 muestra su distribución a lo largo de los ocho PECD analizados:

Figura 7

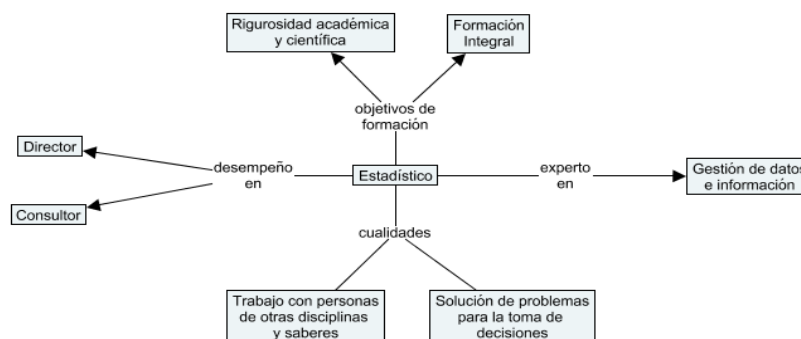
Distribución de la categoría estadístico en los PE



El estadístico se describe a la luz de cuatro grandes aspectos, que también son cuatro categorías importantes dentro del desarrollo de TF: (a) los objetivos de formación, o lo que los PE buscan reflejar en sus objetivos de formación lo que se espera alcanzar dentro de la formación académica; (b) la caracterización que hacen de un estadístico en función de cómo lo definen; (c) elementos relacionados con el perfil ocupacional o desempeño del estadístico y (d) competencias que resaltan los programas. La Figura 8 da cuenta de la relación entre las categorías más importantes con relación al estadístico como profesional del dato:

Figura 8

Relaciones entre las categorías presentes de los PE



5.2.1.1 Objetivos de Formación.

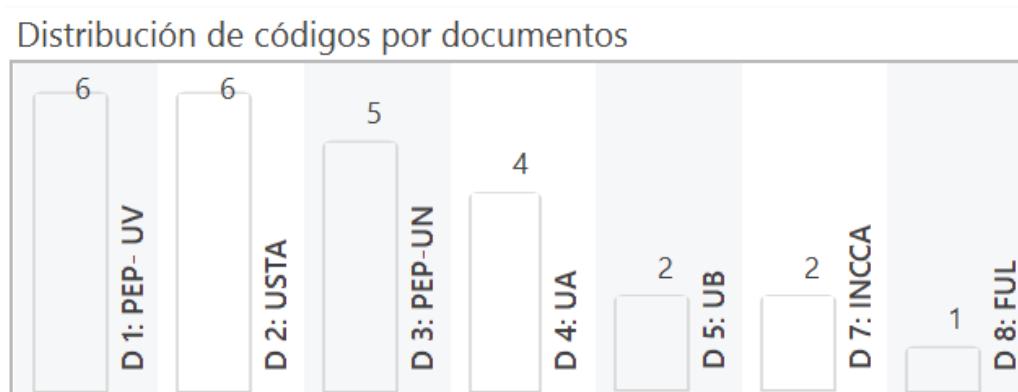
Dentro de los objetivos comunes o mayormente mencionados en los PE se encuentran la rigurosidad académica y científica, y la formación integral. La rigurosidad académica y científica, es entendida como la reunión de características que los programas buscan formar a fin de brindar calidad en el ejercicio disciplinar y ocupacional. Aunque todos los PE persiguen que los estadísticos cuenten con una formación teórica suficiente para aplicar su conocimiento en la solución de problemas reales, cada uno da mayor énfasis en unos u otros aspectos que les parecen más importantes.

Por ejemplo, todos los programas mencionan en mayor o menor medida a las habilidades de investigación como uno de los rasgos propios de la rigurosidad académica y científica (ver Figura 9) y que debe estar al servicio del sector productivo. No obstante, tanto UN como UV visualizan el ejercicio investigativo como la posibilidad de generar conocimiento que permita enriquecer la disciplina,

generando nuevos métodos y herramientas Estadísticas más allá de las que existen en la actualidad, las cuales según UV (2017) son “Análisis Exploratorio de datos, Estimación de parámetros y Pruebas de Hipótesis, Regresión, Muestreo, Diseño de Experimentos, Análisis Multivariado de Datos, Estadística No Paramétrica, Series de Tiempo, y Aprendizaje Estadístico (Statistical Learning)” (p.13).

Figura 9

Distribución de rigurosidad académica y científica en PE



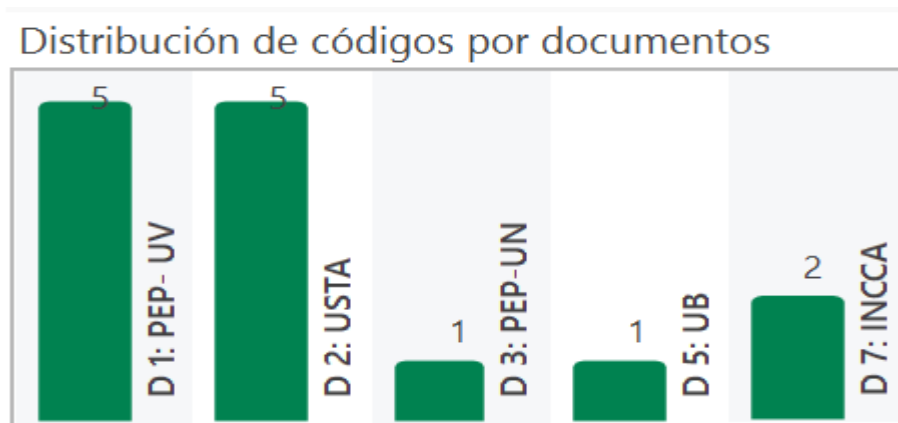
Otro elemento que hace parte de la rigurosidad académica y científica, y del cual no hay unanimidad, es que en algunos PE buscan brindar una especie de valor agregado que está directamente relacionado con la facultad que oferta el programa. Con relación a los programas ofrecidos desde facultades de ingeniería se puede decir que, en el caso de la UV, se menciona repetidamente que el ejercicio estadístico debe estar orientado a la gestión organizacional soportada en la ingeniería; mientras que en ECCI hacen énfasis en las habilidades computacionales. El programa de la USTA pretende dar una formación en aspectos económicos y financieros, componente que también está presente en

INCCA, a pesar de que en esta última el programa de Estadística no es ofertado por una facultad de ciencias económicas.

Otro objetivo que se menciona de forma recurrente, aunque no en todos los PE, es la formación integral (ver Figura 10); sin embargo, cada PE entiende esta de un modo distinto. UV caracteriza la formación integral como el hecho de ser consciente de la realidad y tener responsabilidad ambiental; en INCCA, la formación integral implica el desarrollo de rigor investigativo, el pensamiento crítico y creativo; UB entiende la formación integral desde la ética y principios humanistas. La FUL clasifica la formación integral junto con la necesidad de otros elementos como la ciudadanía crítica, el sentido social, ético, estético y político.

Figura 10

Distribución de formación integral en PE



En algunos casos, la influencia de los valores institucionales es más notoria en la formulación de lo que se entiende como formación integral. La UN es el programa que más menciona la necesidad de formar pensadores estadísticos y que la formación integral es un elemento característico de la formación de estudiantes a nivel institucional, que se incluye dentro de la formación de PE –ver

sección 3.3.2 – al más alto nivel. En el caso de USTA, la formación integral involucra una visión filosófica centrada en el pensamiento de Santo Tomás de Aquino, que implica problematizar la realidad para entender la realidad desde la teoría.

Otros propósitos de formación menos comunes, pero mencionados por algunos PE son los siguientes:

- la capacidad de trabajar con personas de otros saberes y disciplinas (UN, UV, USTA).
- habilidades comunicativas en lengua materna y lengua extranjera (UN).
- capacidad para continuar con éxito su formación académica, traducida en formación posgradual (UN).
- fomentar el pensamiento estadístico (UV, UN).
- otras formas de pensamiento que según UN, interactúan con PE, y que también son mencionadas en otros PE, aunque en forma dispar: lógico, crítico (mencionado también en UV, USTA e INCCA), creativo (mencionado en UV e INCCA), sistémico (mencionado en UV), reflexivo (mencionado en UV), interrogativo, analítico, inductivo y deductivo.

5.2.1.2 ¿Qué es un Estadístico?.

En el PEP de la UN caracterizan al estadístico a partir de una matriz de funciones específicas y estructurales. Por un lado, las específicas encarnan lo relacionado con lo que se espera que haga en su ejercicio profesional, entre las que se incluyen la planeación, la obtención (acopio) de datos, el análisis de datos y el intercambio de información. Merece atención lo mencionado con relación al

análisis de datos –lo cual se ampliará en la sección 5.3–, de donde se puede deducir que es estadístico la referencia más familiar que tiene la sociedad sobre un analista de datos, cuando menciona que

“(…) en la academia, en el medio laboral y en el medio de investigación, se le reconoce como el profesional encargado del análisis de los datos; es de por sí la imagen colectiva que ocupa el sitio más moderado dentro de las imágenes colectivas extremas.” (Universidad Nacional de Colombia, 2015, p. 20).

Entre tanto, las estructurales están más direccionadas con el objetivo máximo de formación deseable: lograr que se vuelva autor de procedimientos, métodos o algoritmos estadísticos, a partir de la generación de conocimiento provocada por modificaciones o adaptaciones a casos particulares, que sean aplicables en forma general. Estas son las que proporcionan medios al estadístico que facilitan su comunicación con otros profesionales de conocimiento y a un nivel multidisciplinar.

Según la UV, el estadístico es un profesional requerido por organizaciones nacionales e internacionales gracias a su experticia en el manejo de bancos de datos y su capacidad para ser asesor en análisis y procesamiento de información. En la USTA se menciona que es un profesional integral en la gestión de datos en función de la toma de decisiones, con capacidades particulares para la investigación; añaden que la difusión de información es un aspecto crucial dentro de sus labores de gestión, pues se espera que sepa compartir su conocimiento

con los demás, y producto de ella, generar espacios reflexivos con relación a los fenómenos y problemas de la realidad.

Desde UA se reconoce que tiene capacidad para solucionar problemas y contribuir al quehacer profesional, elemento presente en otros PECD (UN, UV). En UB se menciona que “[...] es un profesional con la capacidad de idear soluciones a problemas reales mediante el uso de técnicas Estadísticas, de modelar Estadísticamente una situación y tomar decisiones adecuadas para la empresa en condiciones de incertidumbre” (s.f.). UB añade que debe estar en la capacidad de diseñar, gestionar, implementar y administrar proyectos.

En algunos de los PE llaman la atención sobre la responsabilidad social y disciplina académica que debe enfrentar el estadístico, mencionando dos elementos que las resumen: la ética en el manejo de datos y su contribución al crecimiento de la disciplina. Sobre el primer aspecto, UA y FUL resaltan una dimensión ética de la gestión y uso de datos, esto es, que su gestión contribuya al bienestar social que se espera de un estadístico, siendo esta la difusión de pensamiento estadístico entre científicos, profesionales y la sociedad en general.

En cuanto al crecimiento de la disciplina, la UN mencionó que la mayor responsabilidad académica del estadístico está en la difusión del pensamiento estadístico entre científicos, profesionales y la sociedad en general; así como en la apropiación y generación de nuevo conocimiento que enriquezca la disciplina. También, dentro de la visión del programa, la UV declara que el estadístico debe posicionar la Estadística como una herramienta para la investigación y gestión de producción.

Con relación a lo que el Estadístico es capaz de hacer con los datos o con la información, se deduce que el estadístico es un profesional capacitado para toda labor que involucre la gestión de datos e información. Cada universidad mencionó acciones incluidas allí como el diseño y ejecución de planes de recolección y organización (UN, USTA, FUL, ECCI); interpretación (USTA, ECCI), administración (FUL) y análisis (UN, USTA, FUL, ECCI). Algunos programas hacen menciones indirectas al Big-Data, aclarando que estas acciones deben llevarse a escalas más grandes en cuanto a tamaño de datos (ECCI); otras como UV mencionan la necesidad que el estadístico tiene de adaptar sus técnicas y herramientas a situaciones donde el trabajo clásico con datos es insuficiente.

5.2.1.3 Los Roles del Estadístico vs. Facultades Oferentes.

Para materializar lo mencionado con relación a lo que es capaz de hacer un estadístico se han identificado diversos grados de responsabilidad que modelan su ejercicio profesional como director, consultor, analista, asesor, investigador, diseñador, organizador, creador, ejecutor, procesador y evaluador. Contando con que hay programas que no aportan más información con relación a estas posibilidades de perfil ocupacional, una posible forma de triangular la información acerca de los roles del estadístico es comparar sus manifestaciones según el tipo de facultad que ofertan los PE.

En este caso, las mayores fuentes de información provienen de cuatro PE; dos de ellos, que son ofertados por facultades de ciencias y los otros dos, por facultades de ingeniería. Como se verá a continuación, la facultad de origen de

estos PE es un factor que influye en cómo visualizan los diferentes roles del estadístico.

Los programas de la UN e INCCA, ofertados por facultades de ciencias básicas o afines, identifican que el estadístico puede desempeñarse principalmente como consultor o investigador. Sobre el rol de consultor, ambas no se ponen de acuerdo con el ámbito en el que se ejecuta la consultoría; la UN menciona ámbitos como la rama judicial, las entidades financieras, el sector público e instituciones internacionales, mientras que INCCA declara que el ámbito es la industria actuarial, específicamente los seguros y las pensiones. Sobre el rol de investigador, ambas coinciden que realiza investigación teórica pero también en sectores como la salud, las finanzas, el sector agropecuario y el mercadeo.

Por otra parte, los programas de la UV y ECCI, ofertados por facultades de ingeniería, identifican que el estadístico puede desempeñarse como director de oficinas o departamentos de Estadística tanto a nivel privado como público. ECCI profundiza en lo que hace un estadístico como líder en estos espacios: modelación matemática y Estadística orientada a la optimización y toma de decisiones.

5.2.1.4 Los Roles del Estadístico vs. Tiempo de Vigencia.

Tomando como base los mismos roles mencionados en la sección anterior, se establece ahora una triangulación de corte temporal en la que se visualizan los PE de acuerdo con su año de creación o funcionamiento. Por un lado, aquellos con fecha de creación anterior a 2000, o del siglo XX; y por el otro, los posteriores al año 2000, o del siglo XXI. Nuevamente, como se verá a continuación, el tiempo

de vigencia de los PE parece influir en la forma como se manifiestan los roles del estadístico.

De los programas del siglo XX –UN (1958) y UV (1979) –, se identifica que el estadístico es principalmente analista y organizador. Hay consenso en que el estadístico es un analista de datos, en lo que, según UV, se emplean herramientas computacionales existentes y otras que pueda crear en forma independiente. Como organizador, mientras que UN señala que el ámbito de este rol radica en orientar el proceso de planeación en el trabajo con datos, la UV lo ubica en el contexto de oficinas o departamentos de Estadística, proyectos de investigación o estudios con encuestas o propios de la Estadística.

De los programas del siglo XXI –USTA (2005), ECCI (2014), INCCA (2019) y FUL (2020)-, se infiere que el estadístico es director y consultor. Hay acuerdo en que su rol como director lo ejecuta en oficinas o departamentos de Estadística, tal y como con el caso de los PE ofertados desde Ingeniería. En cuanto al rol de consultoría, no hay consensos; USTA e INCCA destacan contextos como la actuaría y la economía; ECCI menciona que es un consultor de datos, usando software especializado; FUL, en ejercicios de investigación.

5.2.1.5 Competencias Distintivas de un Estadístico.

La mayoría de PE menciona dos características con las que buscan dotar al estadístico: el trabajo con profesionales de otras disciplinas y saberes, y la solución de problemas. Ningún PE llega a realizar una conceptualización de tales características, más bien mencionan elementos que están implicados en cada una de ellas o actitudes deseables para que estas características se manifiesten.

Respecto del trabajo con otras disciplinas y saberes, los PE que hacen referencia a esta capacidad la relacionan con la interdisciplinariedad, la multidisciplinariedad o trabajo en equipo, siendo la primera la que más se menciona. La UN, declara que esta característica tiene una dimensión interdisciplinar, que es una cualidad presente en el desarrollo de funciones estratégicas en ejercicios de investigación y en el intercambio de información.

Acerca de esto último, de UN se deduce que este es un ejercicio interdisciplinario, dada la convergencia de una gran variedad de lenguajes - científico, matemático, estadístico, tecnológico y corriente- donde debe garantizarse la correcta interpretación de símbolos y una representación formal de estos lenguajes. Para lograr lo anterior, es necesaria la adquisición de competencias escritas, orales y comunicativas, tanto en lengua materna como extranjera, que permitan al estadístico ejercer de emisor y receptor.

Para ser un buen receptor, se requiere la cualidad de saber escuchar, pues esto le ayudará a aprender, profundizar y asimilar conocimientos implicados en el proceso de interlocución; así mismo, le ayudará a identificar los objetivos, circunstancias, motivos y el problema de una actividad Estadística en particular, y así elegir los procedimientos y métodos adecuados para llevarla a cabo. Por otra parte, para ser buen emisor se requiere del currículo la adquisición de destrezas para dirigirse a todo público, más allá de su formación académica, tanto en español como en inglés, en medios hablados o escritos, siendo claro y preciso, y usando adecuadamente los recursos tecnológicos actuales.

También, la UN resaltó que el trabajo con otras disciplinas y saberes posee una dimensión multidisciplinar. En el perfil de egreso se menciona que el

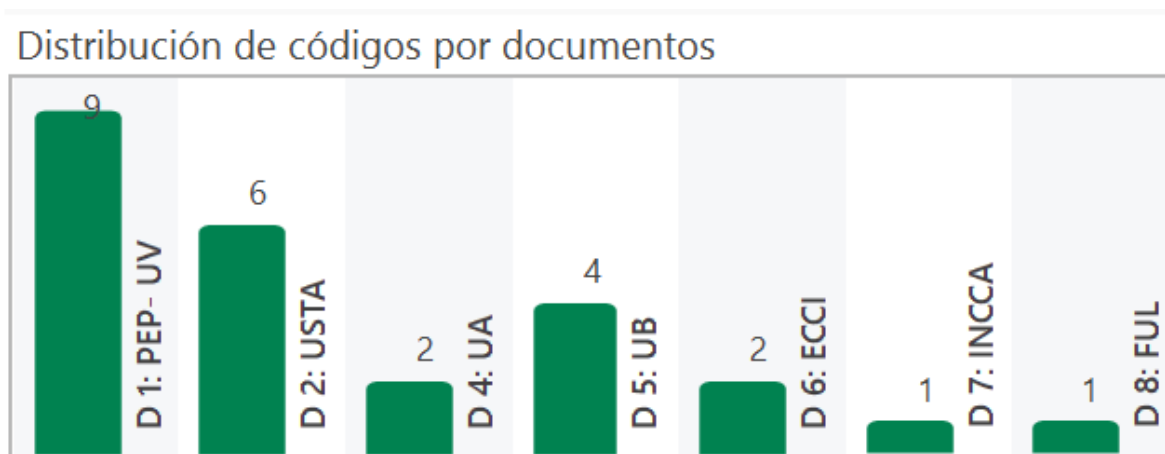
estadístico “tendrá la capacidad de ejercer su profesión de manera multidisciplinaria como canal comunicativo inmanente del grupo a través de la aplicación, adaptación o autoría de procedimientos, métodos o algoritmos estadísticos.” (UN, 2015, p. 10). En otras palabras, su ejercicio de funciones estructurales en el marco de su quehacer profesional es una labor multidisciplinar.

Otros PE que contribuyen en menor medida a caracterizar el trabajo con otras disciplinas y saberes son UV, USTA y UA. La UA menciona que la interdisciplinariedad se pone de manifiesto cuando el estadístico aporta a los esfuerzos de las organizaciones para consolidar datos. UV y USTA coinciden que esta capacidad permite mirar los problemas o fenómenos en forma holística, y que su lugar en equipos de trabajo está orientado –aunque no se limite solo a esto– a la búsqueda y análisis de información, y la elaboración de informes sobre fenómenos para garantizar la toma de decisiones en diferentes contextos.

Justamente, otra cualidad que se menciona en forma recurrente es la solución de problemas y toma de decisiones. Al parecer, hay un consenso que este es el fin de que el estadístico trabaje con personas de otras disciplinas y saberes. La Figura 11 muestra las veces que aparece el concepto en los documentos analizados.

Figura 11

Distribución de solución de problemas en los PE



Pasando a la descripción por cada uno de los PE, aunque la UA menciona explícitamente que el estadístico debe tener la capacidad de solucionar problemas, la mayoría de PE describen lo que se pone en juego al encarar esta tarea. Por ejemplo, la UV habla de una mezcla de visiones -Estadística, social, económica, social, ambiental e integral-, su fundamentación Estadística, valores como el diálogo, la argumentación y la razón; su saber teórico y sus hábitos de trabajo intelectual, en donde se incluye la obtención y el análisis de datos. También añade que la solución de problemas debe estar en función de mejorar la calidad de vida de las personas y mejorar la comprensión de la realidad.

La USTA hace énfasis en la solución de problemas socioeconómicos, por su oferta de formación orientada a elementos propios de las ciencias económicas y finanzas. Con relación a lo que pone en juego para la solución de problemas, además de sus competencias cognoscitivas en el planteamiento de modelos científico-técnicos, emplea competencias afectivas y comunicativas, así como su

aptitud para problematizar la realidad tomando como base la filosofía tomista. Al respecto de esto último menciona que:

La reflexión y el conocimiento permiten hacer preguntas que busquen explicaciones de los fenómenos, con el objeto de encontrar soluciones pertinentes que promuevan el bien de las personas y de la sociedad. Problematizar conlleva a relacionarse de una manera diferente con la teoría, ya que desde este enfoque la teoría es usada para entender la realidad. (USTA, 2017).

UB y ECCI ponen acento en lo que debe saber hacer el estadístico para solucionar problemas, con algunas similitudes y diferencias. Ambas mencionan la recopilación o extracción de datos, así como su análisis. No obstante, mientras la UB habla de la administración de datos mediante el uso de herramientas computacionales y conocimientos en nuevas metodologías Estadísticas, la ECCI habla de organización e interpretación de datos, detección de tendencias, simulación de procesos y optimización de recursos a partir de procesos con grandes volúmenes de datos.

Complementando la descripción de la categoría en el apartado de la toma de decisiones, los PE mencionan condiciones necesarias para que este ejercicio se torne en algo relevante para las empresas y quienes requieran hacerlo. UN mencionó que un requisito para la toma de decisiones estratégicas y urgentes es poder transformar los datos en información, esto es, hacia juicios de valor estratégico, vitales para la toma de decisiones.

La UB menciona en forma general que el conocimiento del estadístico facilita la toma de decisiones, algo que ECCI traduce en la aplicación de modelos estadísticos, matemáticos y de optimización. USTA e INCCA mencionan en forma más explícita a UB y ECCI, que su capacidad en el tratamiento de la información – diseño, recolección, organización, análisis, interpretación y difusión- tiene por fin último que las organizaciones puedan tomar decisiones en forma confiable, eficaz y oportuna. También, FUL añade a lo dicho anteriormente que el contexto de la toma de decisiones está marcado por la presencia de grandes volúmenes de datos.

En resumen, se puede decir que, de acuerdo con lo encontrado en los PE, un estadístico es un experto en la gestión de información. Que desde sus ofertas curriculares pretenden formar principalmente con rigurosidad académica y científica, y formación integral. Que dentro de sus perfiles de ocupación figura el papel de dirigir unidades de Estadística, a pequeña o gran escala, sin importar su procedencia; pero también, el ser consultor en áreas como los seguros, la economía y finanzas, y en cualquier ámbito donde se requiera manejo estadístico. Finalmente, que sus cualidades buscadas son el trabajo con otras disciplinas y saberes, y la solución en problemas para toma de decisiones.

5.2.2 *El Científico de Datos como Profesional del Dato*

En el marco de la formación de profesionales expertos en el tratamiento de datos, el científico de datos se ha convertido en un tipo de perfil emergente, producto de las demandas cada vez más frecuentes, de compañías que requieren personal cualificado para hacer frente a la cantidad de datos que generan. Parte

del alcance del surgimiento de la Ciencia de Datos radica en que ahora se requieren profesionales con habilidades adicionales a las dadas por la Estadística, sobre todo para la gestión de datos no estructurados, que ocupan espacios cada vez más amplios dentro del data set generado por la humanidad.

Tomando en cuenta el análisis hecho con TF sobre los PCD, se procederá a describir cómo estos definen al científico de datos. Posteriormente, y con base en la revisión de información, donde se han podido identificar dos categorías relevantes que tienen que ver con su postura ante los datos, se procederá a describir lo que este tipo de profesional puede hacer con los datos y el para qué trabaja con ellos.

5.2.2.1 ¿Qué es un Científico de Datos?

Entre los cuatro PCD analizados no hay una idea unánime de una conceptualización del científico de datos. El pregrado de la UE ofrece una aproximación cuando afirma que es un profesional capaz de transformar los datos en conocimiento, a partir de una formación basada en la matemática, la Estadística y la computación. De la UDN podría deducirse que logra entender y usar la información que genera la sociedad -en forma de datos- en conocimiento, poniendo en juego habilidades con relación a los datos como su entendimiento, análisis, manejo y síntesis.

Con relación a la información recogida tampoco se observa una lista de características o cualidades comunes en todos los PCD. Nuevamente, la UE hace énfasis en el modo como actúa profesionalmente un científico de datos; el

siguiente listado proporciona características mencionadas y compartidas por al menos otro PCD:

- Son éticos: se espera que el científico de datos asuma una postura ética ante los datos, porque sabe del poder e influencia en la toma de decisiones en el mundo. Este elemento se manifiesta en la FUC, al advertir que el profesional debe adoptar técnicas de uso y protección de información. La PUJ menciona que tienen capacidad para expresar una visión ética, pero no ofrece más detalles sobre este punto.
- Con formación integral: cuentan con competencias científicas, sociales y gerenciales que le permiten hacer frente a problemas complejos. La UDN menciona que las matemáticas, la Estadística y la programación son indispensables en la formación básica de un científico de datos.
- Interdisciplinariedad: son profesionales con habilidad para entender contextos diversos. La PUJ menciona este término por la naturaleza de su oferta, en la que aparecen cuatro facultades que aportan a la formación básica del científico de datos; no obstante, también mencionan que el enfoque de la carrera es multidisciplinar.
- Comunicadores efectivos: dan a conocer sus hallazgos sobre los datos en términos del contexto en el que se encuentran ubicados y de maneras comprensibles para cualquier persona que esté involucrada en él. En la FUC, se menciona esto desde el punto de vista de la adopción de nuevas formas de visualizaciones gráficas de los datos.

- Entienden los retos de las organizaciones: esto es complementado en FUC cuando destacan que un científico de datos se vale del comportamiento de los datos para identificar y diagnosticar sus posibles problemas.

Los demás PCD se concentran en señalar elementos que dan a entender lo que puede hacer a nivel profesional, sin llegar a describir con detalle cada tipo de perfil. En ese orden de ideas, se menciona una gran variedad que un científico de datos podría ocupar, tal y como se aprecia en la Tabla 6, en la que la UE no aporta información alguna. Los campos marcados con Sí representan ideas presentes en el PCD, dentro del perfil de desempeño de un científico de datos.

Tabla 6

Perfiles que el científico de datos puede asumir según PCD

	UDN	PUJ	FUC
Analista de Datos	Sí	Sí	Sí
Asesor	Sí	No	Sí*
Empresario	Sí	No	Sí*
Consultor	Sí	No	Sí**
Analista cuantitativo	Sí	No	No
Docente	Sí	No	No
Científico	Sí	No	No
Investigador	Sí	No	No

	UDN	PUJ	FUC
Arquitecto de Datos	No	Sí	No
Ingeniero de Datos	No	No	Sí
Científico de Datos	No	No	Sí
Gerente de analítica	No	No	Sí

* se menciona el término “gestor de su propia empresa”

**consultor en Analítica, temas de calidad, productividad y mejoramiento continuo.

De la Tabla 6 puede inferirse que el perfil más mencionado en los documentos es el analista de datos y la UDN añade el de analista cuantitativo. La PUJ amplía que un científico de datos puede desempeñarse en actividades de análisis de datos y Estadística con fines sociales. Otras actividades importantes, aunque con especificaciones de los ámbitos de aplicación, serían las de asesor, empresario y consultor. Solo la FUC menciona explícitamente el perfil de científico de datos, mientras que en UDN se menciona que puede desempeñarse como científico, investigador o docente.

5.2.2.2 ¿Qué es Capaz de Hacer con los Datos?.

Una de las categorías más relevantes en el análisis de PCD tiene que ver con la palabra dato, de la cual se muestra un gráfico donde se representa su mención en cada uno de ellos.

Figura 12

Distribución de la categoría dato en los PCD



La información que se encuentra no apunta a una conceptualización del dato, pero va encaminada a caracterizar qué es capaz de hacer un científico de datos con ellos y para qué trabaja con ellos.

Aunque, con relación a lo que se espera que haga un científico de datos con los datos no existe un conjunto de acciones que todos los programas analizados mencionan de la misma manera, se pudo identificar una variedad importante de enunciados presentes en al menos uno de ellos, tal y como se muestra en la Tabla 7.

Tabla 7*Lo que hace un científico de datos con datos según PCD*

¿Qué es capaz de hacer con los datos?	UDN	PUJ	FUC	UE
Desarrollar modelos predictivos en conjuntos de datos estructurados y no estructurados.	No	No	Sí	Sí
Producir informes y visualizaciones para comunicar sus resultados de manera efectiva.	No	No	Sí	Sí
Manejar grandes cantidades de datos.	Sí	No	No	Sí
Hacer análisis de datos exploratorios en conjuntos de datos estructurados y no estructurados.	No	No	No	Sí
Analizar grandes cantidades de datos	Sí	No	No	No
Entender y sintetizar grandes cantidades de datos.	Sí	No	No	No
Generar nuevo conocimiento a partir de los datos.	No	No	No	Sí
Optimizar procesos y encontrar patrones	No	No	Sí	No
Medir la calidad de información que se recoge, cura según su pertinencia.	No	No	No	Sí
Gestionar datos de manera estratégica.	No	No	Sí	No

Hay tres acciones que se podrían destacar de la lista anterior: la predicción, la visualización y el manejo. La predicción es materializada como la capacidad de desarrollar modelos predictivos independiente de los datos si son estructurados o no; al respecto, FUC profundiza que las herramientas para lograrlo provienen de las tecnologías basadas en sistemas de información en la nube, Machine Learning, Inteligencia Artificial y Analítica de Datos. La visualización está relacionada con la producción de informes que ayuden a comunicar las ideas relacionadas con los datos de manera efectiva, para lo cual debe valerse de nuevos métodos gráficos y, en definitiva, del uso de herramientas Estadísticas y tecnológicas recientes. Sobre el manejo, se hace énfasis en que debe hacerse con grandes cantidades de datos.

Otras acciones que emergen en menor medida de la tabla anterior están relacionadas con el análisis de datos. Al respecto, la UDN hace énfasis en la necesidad de hacerlo con grandes cantidades de datos, además de añadir acciones como el entendimiento, el manejo y la síntesis de estos. Así mismo, la UE especifica en un tipo de análisis como el exploratorio, tanto en datos estructurados como en no estructurados. Otros enunciados que están menos presentes son la generación de nuevo conocimiento a partir de los datos, la gestión estratégica, el control de calidad, la sistematización con minería de datos, la optimización de procesos y la deducción de patrones.

5.2.2.3 ¿Para qué Trabaja con Datos?

A pesar de lo anterior, parece existir un común acuerdo en que los fines últimos de lo que un científico de datos hace con los datos son la toma de

decisiones y la solución de problemas; estas últimas, dos categorías que emergen de los datos por su continua mención en los documentos analizados, tal y como se visualiza en las Ilustraciones 12 (toma de decisiones) y 13 (solución de problemas).

Figura 13

Distribución de la categoría toma de decisiones en los PCD

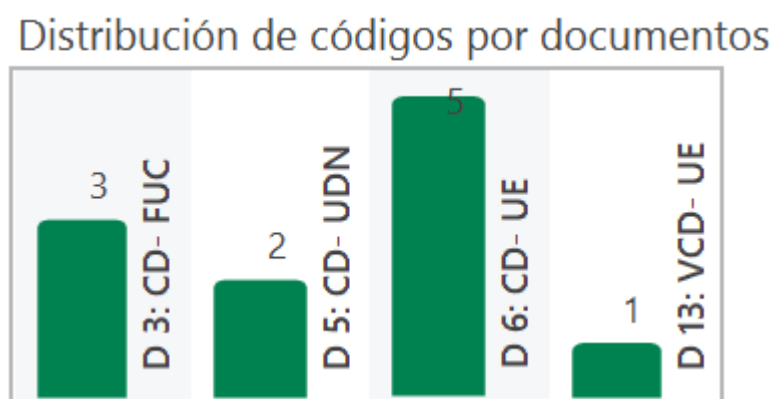
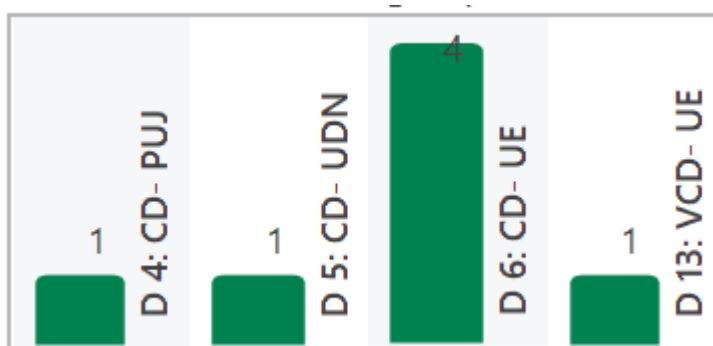


Figura 14

Distribución de la categoría solución de problemas en los PCD



Al respecto, la Tabla 8 destaca una serie de enunciados que encarnan para qué un científico de datos trabaja con datos.

Tabla 8*El para qué del trabajo con datos según PCD*

¿Para qué se trabaja con datos?	UDN	PUJ	FUC	UE
Tomar decisiones a partir de los datos.	Sí	No	Sí	Sí
Formular propuestas de innovación con base en su comportamiento.	No	No	Sí	No
Implementar propuestas de innovación con base en su comportamiento.	No	Sí	Sí	No
Resolver problemas a partir de los datos	Sí	No	No	Sí

El elemento que parece ser más importante en la mayoría de los programas analizados es la toma de decisiones. Los programas analizados no conceptualizan cómo un científico de datos toma decisiones a partir de los datos, hacen más énfasis en las herramientas de las que dispone para lograrlo. Todos los programas que mencionan la toma de decisiones como uno de los fines del trabajo con datos declaran que la inteligencia artificial es una herramienta para lograrlo. No obstante, de la UDN se deducen otras como la inteligencia de datos, la optimización, el análisis estadístico y otras que no menciona en forma explícita; la FUC habla de las tecnologías basadas en sistemas de información y en la nube, Machine Learning y Analítica de Datos. La UE menciona herramientas de modelación, predicción y automatización.

En cuanto a la solución de problemas, nuevamente no hay definiciones de cómo el científico de datos lo hace, sino en los conocimientos de base que

requiere para lograrlo. De UDN y UE se puede inferir que las mismas herramientas mencionadas para la toma de decisiones también intervienen en la solución de problemas, algo que la FUC no ratifica en su información. Solamente UDN y UE mencionan explícitamente la solución de problemas a partir de los datos y ambas coinciden que para la solución de problemas es fundamental una fuerte formación matemática, Estadística y computacional; sin embargo, mientras que UDN añade la formación en gestión y comunicaciones, la UE añade habilidades gerenciales e interdisciplinarias.

Adicionalmente, hay una serie de enunciados que podrían asociarse con la solución de problemas como formular propuestas de innovación, implementar -o desarrollar- con base en el comportamiento de los datos; de hecho, FUC especifica que estas se desarrollan a partir del análisis de datos y que son procesos que deben estar encaminados a la mejora de problemas ambientales, sociales y culturales. Finalmente, otras acciones vinculadas son la identificación y diagnóstico de problemáticas organizacionales (UDN), el desarrollo de sistemas y modelos para la solución de problemas complejos (UE), y la puesta en marcha de casos de prueba para la revisión y validación de propuestas a soluciones de problemas (UE).

5.2.3 ¿Qué es un Ingeniero en Ciencia de Datos?

De la información sobre los ICD se levantó la siguiente tabla que clasifica los diversos roles en los que puede desempeñarse un ingeniero en Ciencia de Datos. Justamente, una de las categorías más relevantes con relación a este

profesional de los datos es la de analista, siendo este el tipo de rol más identificado por las universidades oferentes de ICD.

Tabla 9

Roles identificados por los ICD

Rol	GJ	PGC	PUB	CUIB
Analista	Datos. Riesgos. Inversiones. Equipos multidisciplinarios	Proyectos de investigación	Negocios. Big-Data. Calidad de datos. Ciencia de Datos.	Datos.
Director	No lo menciona	Analítica. Inteligencia Empresarial.	Analítica. Gestión de información. Gobierno de Datos.	Proyectos de Big-Data. Infraestructuras de Big-Data. Analítica avanzada de negocio. Chief data officer.
Ingeniero	Datos.	No lo menciona	Datos.	No lo menciona
Científico	Datos.	Datos.	No lo menciona	No lo menciona
Arquitecto	Datos.	No lo menciona	No lo menciona	Soluciones Big- Data. Ciencia de Datos.
Administrador	Datos.	No lo menciona	No lo menciona	No lo menciona
Investigador	Negocios.	Proyectos de investigación.	No lo menciona	No lo menciona

Rol	GJ	PGC	PUB	CUIB
Consultor	No lo menciona	Proyectos de investigación.	No lo menciona	Consultor externo.
Protector	Datos.	No lo menciona	No lo menciona	No lo menciona
Gestor	Analítica. IT.	No lo menciona	No lo menciona	No lo menciona
Coordinador	Estadística. IT.	No lo menciona	No lo menciona	No lo menciona
Minero	No lo menciona	No lo menciona	Datos.	No lo menciona
Desarrollador	No lo menciona	No lo menciona	Ciencia de Datos. IA.	No lo menciona

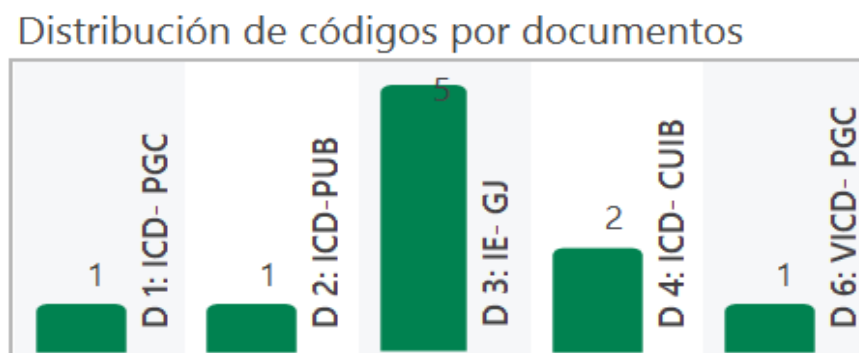
Aunque no es evidente el rol con el que más se vinculan los ICD, si surgen dos tipos de ocupaciones mencionadas en la mayoría de las ocasiones: (a) el de analista de datos, o por lo menos, de acciones que deben evaluarse de ellos como la calidad y el gobierno (b) el de director en analítica. En un segundo rango de roles pueden identificarse el de analista de negocios (o condiciones relacionadas con ellos como los riesgos e inversiones); consultor; e ingeniero, científico y arquitecto de datos. Otros roles menos importantes tienen que ver con la minería de datos, actividades de investigación, gestión y coordinación.

Con relación a categorías importantes dentro del análisis hecho con TF sobresale la solución de problemas, aunque, como en los casos precedentes, no se brindan definiciones ni ideas, sino que se pone énfasis en las herramientas y tipos de problemas que enfrentará. Con relación a las herramientas, salen a la luz

adjetivos como predicción, simulación, optimización, implementación de tecnología, competencias prácticas, teóricas, comunicativas y la ética. Frente a los tipos de problemas que podría encontrar pueden ubicarse contextos como la industria, la academia y la ciencia. La Figura 15 da cuenta de cómo la categoría solución de problemas es nombrada en cada ICD.

Figura 15

Distribución de solución de problemas en los ICD



De igual manera, el fin último de la solución de problemas es la toma de decisiones, categoría que no resultó tan relevante como en el caso de PE y PCD.

5.3 El Análisis de Datos en PECD

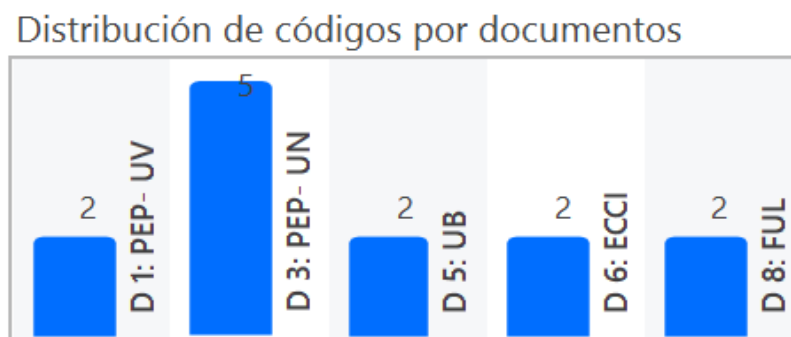
Para terminar con el capítulo de resultados obtenidos a partir del análisis de los PECD, se busca identificar si el análisis de datos es un aspecto medular o no en el quehacer del profesional experto en datos.

5.3.1 Según los Programas de Estadística

La categoría análisis de datos resulta ser mencionada en la mayoría de los PE (5 de los 8 PE), La Figura 16 muestra la distribución de esta categoría por cada PE, siendo el de la UN quien más la nombra.

Figura 16

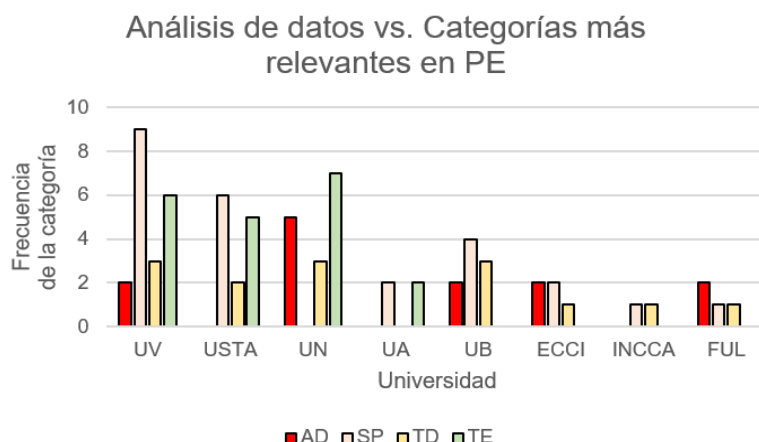
Distribución de la categoría análisis de datos en los PECD



Pese a lo anterior, no alcanza un lugar tan relevante como las descritas en la sección 5.2.1: la toma de decisiones (TD), la solución de problemas (SP) o el trabajo con otras disciplinas y saberes (TE). Así mismo, con excepción de ECCI y FUL, no hay ningún PE en el que el análisis de datos (AD) sea la categoría más encontrada en los PE si se compara con las que son principales. La Figura 17 resume lo anterior.

Figura 17

Comparativo de análisis de datos frente a categorías en PE



El programa que más información posee con relación a esta categoría es el de UN. El PEP lo incluye como parte medular del ejercicio profesional y ofrece una descripción detallada del mismo. Luego de situar al estadístico como el analista de datos por excelencia (ver sección 5.2.1.2), justifica esta función específica –según la estructura curricular del programa– en la necesidad de convertir datos en información o juicios de corte estratégico. Así mismo, declara que es “el medio que permite la descripción en detalle y se convierte en una fuente que nutre la toma de decisiones apremiantes y estratégicas” (UN, 2015, p. 20).

Adicionalmente, listan una serie de funciones detalladas o acciones que instrumentalizan el análisis de datos en el estadístico de la UN: (a) la articulación de bases de datos existentes para consolidar uno o varios conjuntos de datos con criterios de relevancia y pertinencia; (b) la depuración de bases de datos con base en lo establecido en lo planeado para obtenerlos y procurando su confiabilidad; (c) el tratamiento de outliers y datos faltantes para estudiar sus efectos; (d) la elección

de procedimientos adecuados de análisis en función de los objetivos de un proyecto o investigación, incluyendo la elección de software adecuado; (e) la elección de métodos compatibles con la estructura de los datos y (f) la evaluación de procedimientos aplicados para el análisis de datos para ratificarlos o buscar alternativas.

La UN también sugiere que el análisis de datos es una función posterior a otras dos específicas como la planeación y el acopio de datos. De los demás PE que la nombran categoría, se puede ratificar esta misma idea. Por ejemplo, en UV parece ser una acción que está precedida por la comprensión, planeación y obtención de datos. La UB la ubica luego de la captura y el procesamiento de datos. La ECCI, posterior a la recopilación, organización e interpretación de datos. En FUL se menciona posterior a la extracción y previo a la administración de datos.

Así mismo, desde las herramientas Estadísticas que la UV identifica como fundamentales en el ejercicio profesional, cabe el análisis exploratorio de datos y el análisis multivariado. Una revisión del plan de estudios de la UN sugiere varios tipos de análisis de datos: análisis de datos categóricos, análisis de datos longitudinales y análisis multivariado.

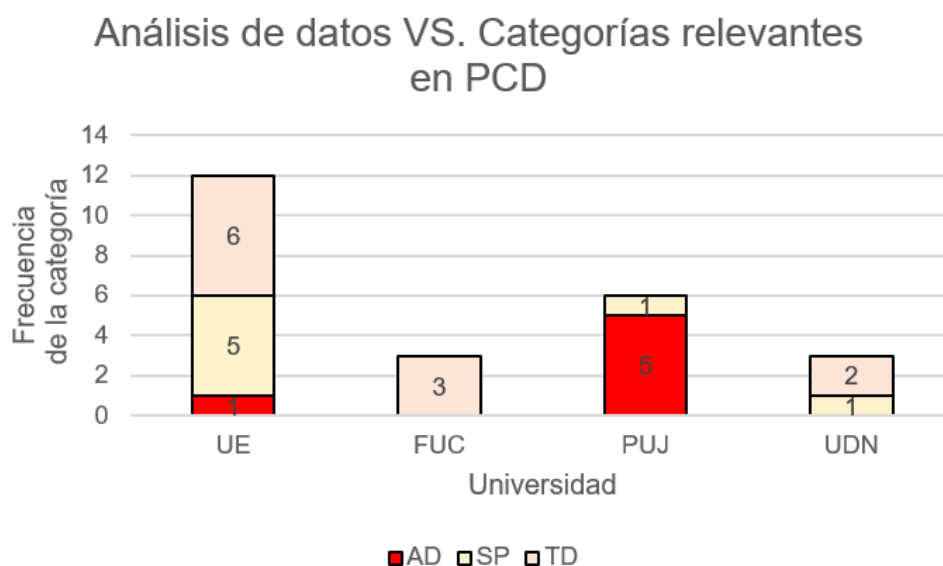
Finalmente, en un apartado de aspectos variados con relación al análisis de datos, la UB nombra al análisis de datos como uno de los aspectos de estudio de la Estadística como disciplina y FUL hace referencia a que el conocimiento en nuevas metodologías Estadísticas y formación computacional avanzada son herramientas para el análisis de datos.

5.3.2 Según los Programas de Ciencias de Datos

Un procedimiento similar al mostrado en PE muestra nuevamente, que el análisis de datos es menos relevante que la solución de problemas y la toma de decisiones, dos de las categorías más relevantes en la caracterización de un científico de datos (ver sección 5.2.2.1). A continuación, la Figura 18 representa esto; también se observa que, salvo PUJ, no hay ningún PCD en el que el análisis de datos (AD) sea la categoría más encontrada si se compara con las que son principales en los PCD.

Figura 18

Comparativo de análisis de datos frente a categorías en los PCD



A diferencia de los programas de PE, no hay conceptualizaciones detalladas que puedan encontrarse. PUJ menciona que es una capacidad que el científico de datos desarrolla para ponerla en función de la solución innovadora de problemas en la ciencia, las empresas y la sociedad. La UDN afirma algo similar,

pero se refiere al análisis de información en lugar de análisis de datos. La UE sugiere que las herramientas computacionales y los avances estadísticos recientes son condiciones para el análisis de datos. Finalmente, los programas de ICD no hacen mención alguna del análisis de datos. Tal vez, el término más afín tenga que ver con la analítica de datos, algo que tampoco llegan a caracterizar a fondo.

6. Discusión y Conclusiones

Luego de recopilar y analizar la información referente a los PECD, se listan reflexiones y discusiones en torno a los tres objetivos específicos de la investigación. La primera sección propone consolidar lo encontrado con relación a la forma como se revela la Sociedad del Dato en los PECD

6.1. Acerca de las Manifestaciones de la Sociedad del Dato

En líneas generales, hay tres aspectos que se manifiestan desde los PECD con relación a la Sociedad del Dato: la producción masiva de datos, la dimensión ética del uso de los datos y el reconocimiento de la Estadística o Ciencia de Datos como disciplinas transversales a la mayoría de los campos de conocimiento o al servicio de la solución de problemas reales. A continuación, el desarrollo de estas ideas.

6.1.1 *Big-Data: ¿Sólo Volúmenes Masivos de Datos?*

Parece haber un acuerdo en que los tres tipos de PECD reconocen la presencia de Big-Data, explícita o implícitamente como un fenómeno moldeado por la generación masiva de datos. Con relación a los PE, quizá por el escaso conocimiento que existe en el medio colombiano con relación a temas de Big-Data, tema del que se ha hablado con mayor fuerza en la última década, las propuestas de pregrados más recientes son más intencionales en reconocer que la producción masiva de datos trastoca la labor del profesional en Estadística.

Con relación a los PCD e ICD, uno de los argumentos que más brindan con relación a la relevancia de sus programas es la importancia que el mundo ha venido asignando a los datos en el desarrollo de la humanidad. Insisten con frecuencia que los grandes volúmenes de datos han supuesto nuevos retos para la humanidad, entre ellos, su procesamiento y convertirlos en información que resulte en nuevo conocimiento.

En realidad, el término Big-Data solo es mencionado dos veces en los documentos relacionados con los PE, y la característica que más resaltan del fenómeno tiene que ver con los grandes volúmenes de datos. A diferencia de los PE, la categoría dato está mucho más presente en los PCD, en donde a partir de la información encontrada en ellos se pudo identificar por qué y para qué se trabaja con datos -ver secciones 5.2.2.2 y 5.2.2.3-. Por ser el Big-Data un fenómeno que tiene una relación más estrecha con lo que hace un científico de datos, es de esperar que la mención de este término haya sido más frecuente en los PCD que en los PE.

Aunque en menor medida, otra característica que reconocen los PECD con relación al Big-Data es la variedad en la que vienen los datos. En el caso de los PE, solo uno de los programas llega a reconocer que el estadístico está inmerso en situaciones que involucran datos cualitativos o textuales. En el caso de los PCD llaman la atención que casi cualquier tipo de actividad en la que exista mediación de la web constituye un entorno para la generación de datos, incluyendo grabaciones de video, fotos, imágenes, archivos no textuales, entre otros, nociones más relacionadas con los datos no estructurados.

Si bien, la característica más llamativa del fenómeno Big-Data puede ser el volumen –cada vez más grande– de datos, las revisiones teóricas del concepto presentadas en esta investigación, revelan que Big-Data se trata de mucho más que producción masiva; habría que incluir en este fenómeno la velocidad de generación de datos y el componente de incertidumbre que siempre ronda la obtención de ellos (Camargo-Vega, Camargo-Ortega y Joyanes-Aguilar, 2015; Paredes-Moreno, 2015; Vásquez, 2020), lo que según GAISE (2016) supone retos para formar competencias en análisis de datos a nivel mundial y en cualquier nivel educativo. En ese orden de ideas, los PECD ofrecen una visión limitada de lo que entienden por Big-Data. Formar pensamiento estadístico en los futuros profesionales del dato se vuelve fundamental en las propuestas curriculares porque los prepara para modelar la variabilidad presente en los datos (Moore, 1997; Hoerl y Snee, 2010; GAISE, 2016).

6.1.2 *Ética en el Trabajo con Datos: ¿Justicia o Beneficios?*

La mención de la ética dentro del ejercicio profesional es tratada como un aspecto vital en el desempeño del estadístico, pero solo desde el punto de vista de garantizar calidad. Los PE y PCD no ahorran palabras en señalarla como parte fundamental del proceso de formación profesional, que es un sello distintivo de los trabajos con datos bien hechos desde un punto de vista técnico y que ser ético implica perseguir el bien común o el mejoramiento de la calidad de vida. Entre las razones para la ética que declaran los PCD, se menciona que el científico de datos, al ser consciente del poder de los datos en la toma de decisiones que afectan la humanidad, asume una postura ética. Algo que llama la atención es que

en ningún ICD se menciona la ética como parte de la formación del ingeniero en Ciencia de Datos.

No obstante, tal y como O'Neill (2016) lo comentaba, aún no se ha reflexionado lo suficiente sobre el impacto que tienen los datos en la vida de las personas. En ese orden de ideas, el ejercicio de la ética no puede limitarse solamente a realizar bien el trabajo. Aunque hay una gran multitud de voceros y dataístas en el mundo que pregonan el poder de los datos, es necesario que se revise cómo hacer de su tratamiento algo no tan costoso en términos de las luchas más terribles de la humanidad: combatir la desigualdad, cerrar brechas discriminatorias, procurar justicia para los sectores más marginados, entre otras cosas.

Si el dato es el nuevo petróleo y su extracción es necesaria para la existencia, se debe crear conciencia de la necesidad de eliminar efectos colaterales que causen el menor perjuicio posible a la humanidad en términos de su privacidad y el ejercicio de sus derechos. O'Neill (2016) sugiere que detrás de todo trabajo con datos sobre personas, debe haber información que le indique a la gente cómo serán tratados sus datos, pero también los potenciales efectos y riesgos que esto podría tener. En ese orden de ideas, no está mal que, tanto los que trabajan con datos como las empresas o centros de investigación que los poseen, busquen los beneficios que supone la generación de conocimiento; pero no se puede lograr menoscabando la justicia que debe primar para las personas que los generan.

El uso ilimitado e irrestricto de datos, aún con buenas intenciones, es motivo de un debate al que no pueden escapar los PECD y del que deben generar

más certezas que dudas, después de todo, no es exagerado pensar que, como dijo Harari (2017), quien tiene los datos tiene el futuro.

6.1.3 Estadística y Ciencia de Datos: ¿Discontinuidad Evitable?

Ya sea que se trate de los PE con la Estadística, o los PCD e ICD con la Ciencia de Datos, cada tipo de programa reconoce y justifica la importancia de su disciplina de base. La Ciencia de Datos, como campo que surgió de controversias en el seno de la comunidad Estadística a mediados del siglo pasado, se ha manifestado no solo como un campo emergente, sino que pretende presentarse como una disciplina que brilla con luz propia, al margen de lo que la Estadística pueda ofrecer.

Si se quisiera brindar una definición de la Estadística a partir de los PE, se puede proponer que es una disciplina orientada al estudio de los procesos relacionados con la gestión de datos e información, que incluyen la recolección de datos, su análisis y la interpretación a la luz del contexto en el que están ubicados. Los PE reconocen su disciplina de base como un campo de conocimiento cuya influencia es cada vez mayor en el desarrollo de otras áreas de conocimiento; con un altísimo grado de confiabilidad puede asegurarse que no hay contextos en los que la Estadística tenga algún tipo de uso, pero aparecen escenarios que se mencionan como la salud, la economía y la investigación.

Ahora bien, si los PE reconocen un mundo permeado completamente por los datos entonces resulta preocupante que no asuman posturas con relación a la Ciencia de Datos, o el lugar que la Estadística tiene con relación a ella. Si bien, aún persisten voces que las asumen como sinónimas (Moore, 1991; Broman 2013,

citado por Fienberg, 2014), y que en el caso colombiano la línea de tiempo de los PE (1958- 2020) es mucho mayor que los PCD o ICD (2019- 2022), la mayoría de los sectores han considerado a la Ciencia de Datos como respuesta a necesidades que consideran, no fueron solventadas completamente desde la Estadística.

Si desde la revisión documental se ha logrado establecer que ningún PE reconoce la Ciencia de Datos como campo emergente, en el caso de los PCD e ICD las cosas son ligeramente distintas. Si bien, no es de su interés realizar una conceptualización de la Estadística como en el caso de los PE, algunos afirman que es un componente básico en la formación de científicos de datos. Lo más cercano a una descripción de la Estadística sucede cuando en el video de la UDN se advierte de la controversia entre estadísticos y matemáticos sobre la necesidad de generar hipótesis para estudiar una población de interés.

En todo caso, tanto los PCD como los ICD se concentran en brindar reconocen en forma parcial la Estadística, aunque no la dejan de mencionar como algo que hace parte de la Ciencia de Datos. Con relación a lo que se menciona sobre la Ciencia de Datos, y en aras de brindar una definición desde los PCD e ICD, se puede afirmar que es una disciplina cuyo interés se centra en el tratamiento de datos orientada a la generación de nuevo conocimiento, a partir de la matemática, la Estadística y la computación.

Algo que obstaculiza el reconocimiento de la Estadística en los PE puede atribuirse a que no hay un consenso sobre el contexto en el que la visualizan. La propuesta de UB sugiere, al igual que Yañez (2000) que la Estadística es una disciplina del siglo XX, amparado por el desarrollo de técnicas inferenciales y el

poder del razonamiento inductivo en el procesamiento de datos y el estudio de la incertidumbre; pero la información de FUL la ubica en el contexto de la Revolución 4.0, término que algunos autores sitúan a mediados del siglo pasado (Schwab, 2016) y otros en pleno siglo XXI (Cañón et. al., 2017), en donde la humanidad ha dado muestras de su poder para superar las dificultades de no poder procesar datos de tipo poblacional y estudiarlos directamente.

Del mismo modo que ocurre con los PE, existen diversas versiones de lo que consideran el contexto de desempeño del científico de datos o el ingeniero en Ciencia de Datos. La propuesta de la PUJ reconoce un contexto similar al reconocido por la FUL, la cuarta revolución industrial, un escenario que identifican principalmente con la continua generación de datos. Entre tanto, GJ ubica al ingeniero de datos en el contexto de la revolución 5.0. acudiendo a fenómenos propios de la cuarta revolución industrial (Ciencia de Datos, transformación tecnológica y digital). En realidad, este escenario está asociado a la sociedad 5.0., término mencionado en Japón a raíz del quinto plan básico de Ciencia, Tecnología e Innovación en 2015, y que implica retos diferentes porque exige formación de personas con suficiente pensamiento crítico como para redefinir la relación hombre-máquina (Gallardo, 2020).

Aunque lo anterior no sea suficientemente contradictorio, asumir contextos diferentes en la formación del estadístico, del científico de datos, y en general de cualquier profesional, puede ofrecer visiones desactualizadas y poco pertinentes de la disciplina de base. Desde la Educación Estadística se ha empezado a reconocer que la Estadística del siglo pasado es totalmente distinta a la de la actualidad (GAISE, 2016; Utts, 2016; Blanco, 2018), en gran parte por la

naturaleza cambiante de los datos que se producen ahora (Gould, 2010) y la emergencia de la Ciencia de Datos, lugar en el que la Estadística tiene un papel importante (ASA, 2015).

Con relación a la formación en Ciencia de Datos en el caso colombiano, se debe reconocer si realmente estamos ubicados en un escenario como la revolución 4.0 o la 5.0 para saber hacia dónde vamos. Se hace necesaria mayor investigación educativa que instale la enseñanza de la Ciencia de Datos como un nuevo campo de investigación, donde se reflexione si el verdadero reto está en formar científicos desde lo técnico, con altos estándares científicos y computacionales, o si se requieren de otra clase de habilidades -incluso blandas- que les permitan superar la barrera de la dependencia hacia la red, y hacer de la relación hombre-máquina algo más negociado que destinado a la conexión perpetua sin relevancia para la existencia humana.

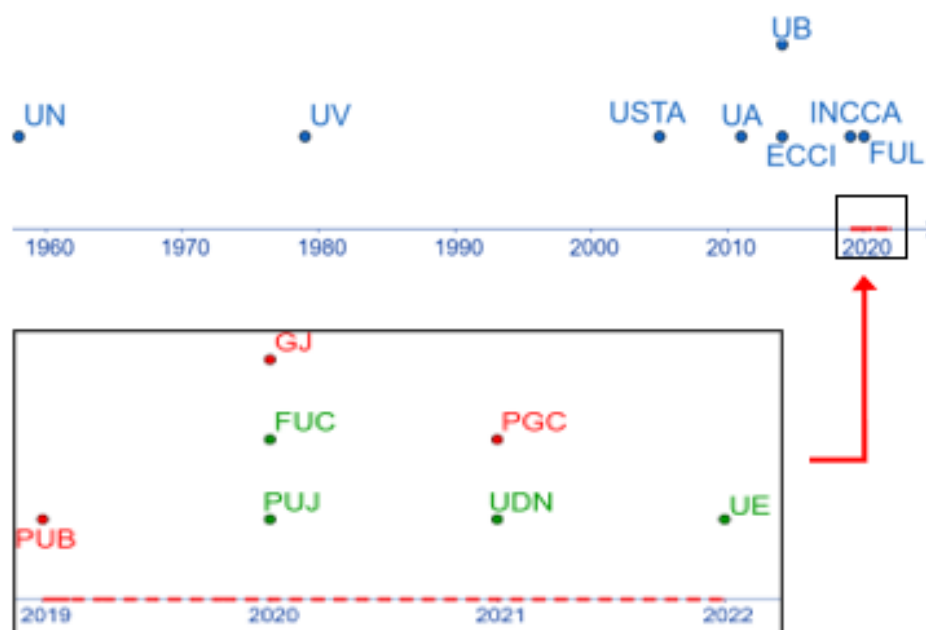
6.2. Acerca del Profesional del Dato

Con el fin de consolidar lo encontrado en los PECD sobre el profesional del dato, se optará por profundizar en los hallazgos encontrados sobre el estadístico, el científico y el ingeniero en Ciencia de Datos. Advirtiendo que no hay unanimidad de criterios con relación a lo que se entiende por un estadístico, y con el ánimo de brindar una definición de lo que es un estadístico a partir de los PE y las categorías encontradas en la TF, habría que decir que es un profesional experto en gestión de datos e información, con formación integral y rigurosidad académica y científica, y habilidades para trabajar con personas de otras disciplinas y saberes en la solución de problemas para la toma de decisiones.

Con relación a los diversos roles que puede asumir un estadístico, y en virtud que la línea de tiempo de los PE es mucho mayor que la de los otros tipos de PECD (ver gráficos siguientes), hubo más posibilidades de formar subgrupos de PE de acuerdo con la facultad que los oferta y con el año de creación o funcionamiento. La Figura 19 muestra líneas de tiempo de los PE (línea de color azul), y de los PCD e ICD (línea de color rojo punteado). Los puntos de color rojo representan los PCD mientras que los de verde representan los de ICD. Cabe recordar que el programa de CUIB no muestra el año de comienzo de su pregrado.

Figura 19

Líneas de tiempo de los PE, los PCD e ICD



Con el análisis mostrado en la sección 5.2. pudo comprobarse que la facultad de origen de los PE es un factor que incide en la manera como se perciben los roles del estadístico. Las facultades de ciencias privilegian habilidades orientadas a la investigación y la consultoría; en cambio, las de ingeniería, habilidades dirigenciales.

Así mismo, la formación de estadísticos en Colombia también está permeada con la tradición o vigencia del PE. Resulta sorprendente que los programas del siglo XX identifiquen al estadístico como un analista de datos y que esto no ocurra con los del siglo XXI, siendo este último un tiempo en el que se reclama con más ímpetu la necesidad de fomentar análisis de datos a niveles incluso más complejos que en el siglo XX, donde proliferaban las bases de datos estructuradas.

Por otro lado, con relación al científico de datos y el ingeniero en Ciencia de Datos, cuesta trabajo identificar elementos comunes y ante la divergencia de puntos de vista sobre lo que es un científico de datos, habría que decir que es un profesional que interesado en transformar grandes volúmenes de datos en conocimiento útil para la resolución de problemas y la toma de decisiones. Por otra parte, da la sensación de que desde los ICD se busca en realidad formar científicos de datos, aunque con roles ligeramente distintos a los que proponen los PCD.

Al igual que la visión de estadístico que proponen los PE del siglo XX, la mayoría de los PCD coinciden que el rol por excelencia del científico de datos es ser analista de datos, rol que también fue identificado en algunos de los ICD. Con relación a estos últimos, también se destacan que sus profesionales pueden

desempeñarse como directores de analítica de datos, lo que cae dentro del dominio del análisis de datos porque incluye herramientas informáticas para encontrar tendencias y elaborar conclusiones de la información que contienen los datos.

En resumen, aunque existan diferencias marcadas entre las configuraciones de roles de los tres tipos de perfiles relacionados con el profesional del dato, puede constatarse que, hasta cierto punto hay visiones compartidas de que este es un analista de datos. Habrá que revisar entonces si el lugar que ocupa el análisis de datos, en coherencia con lo que se ha dicho, es modular o no en su formación profesional.

6.3. Acerca del Análisis de Datos

Advirtiendo la complicación de encontrar puntos comunes en lo que se entiende en los PECD por análisis de datos, podría decirse que es la acción que ejecuta el profesional del dato, valiéndose de su formación Estadística y computacional, para describir e interpretar grandes volúmenes de datos, que convertidos en información puesta al servicio de la toma de decisiones estratégicas y valiosas para la humanidad.

Si se ha identificado que el rol más mencionado en los estadísticos formados de PE del siglo XX, PCD e ICD es el de analista de datos, cabe preguntarse, ¿es el análisis de datos una competencia modular en los PECD? Como ya se había dicho en secciones anteriores -ver sección 5.3- y con excepción del programa de la UN, la revisión de categorías hecho con TF muestra que el

análisis de datos no es una categoría central, con lo que se puede afirmar que su lugar es secundario con relación a otras habilidades o competencias.

Comparativamente, parece que los PE dan un lugar más importante al análisis de datos que lo mencionado en los PCD e ICD. De hecho, el programa de la UN advierte que existe la visión según la cual, la labor del estadístico se centra prácticamente en el análisis de datos. En los demás PE se encuentran acciones que, mezcladas con el análisis de datos, configuran el proceso de gestión de datos e información que hace un estadístico como lo son la recolección y la interpretación a la luz de un contexto.

Entre tanto, cuesta encontrar una idea clara de cómo el científico de datos lo hace, y la mención que hacen los ICD de la categoría es prácticamente nula. De hecho, llama la atención que en su código profesional cataloguen a la Estadística como el escenario donde se hace análisis de datos, y que el científico de datos se identifique de liberar y encontrarle significado a los datos en bruto. Probablemente, si se asume la definición brindada de Ciencia de Datos –ver sección 6.1.3– y que la Estadística incluye al análisis de datos como una de sus acciones de estudio, se podría ver que el científico de datos no puede hacer análisis de datos prescindiendo de la Estadística.

6.4. Conclusiones

Hasta este punto de la investigación puede verificarse que, más allá del nombre que quiera otorgarse a la sociedad (de la información, 4.0, 5.0, etc.), no es descabellado asumir un lente data-céntrico para visualizarla, al punto de llamarla Sociedad del Dato. Hay antecedentes históricos y sucesos recientes que ratifican

una creciente necesidad de emplearlo como forma de representación de casi cualquier fenómeno de interés en la humanidad; áreas de conocimiento como la matemática, la Estadística y la informática, que han interactuado para dar lugar a disciplinas como la cibernética, la neurociencia y la Ciencia de Datos como campos que nutren una visión dataísta del mundo.

Si bien, el uso de los datos representa una oportunidad para la humanidad de mejorar la calidad de vida y brinda nuevas posibilidades de comunicación entre campos de conocimiento, fomentando la interdisciplinariedad y generando inquietudes que motivan a la reflexión académica y la investigación. No obstante, aún la humanidad no ha medido los límites y peligros que supone admitir una visión donde el centro de la existencia sea el dato. En ese sentido, es necesario que desde los escenarios educativos se formen a las personas para asumir posturas críticas, no solamente para tomar decisiones con relación a la información derivada de los datos que se analizan, sino para tomar conciencia de los abusos y malentendidos que pueden ocasionar en la existencia de las personas.

Formar a las personas en competencias como la alfabetización Estadística y el pensamiento estadístico es un buen comienzo para crear una sociedad Estadísticamente culta; sin embargo, quizás sea el momento de procurar un mundo data-alfabeto, algo en lo que la Ciencia de Datos podría aportar, gracias a su intención de aprender a generar conocimiento desde los datos y desde el cual se puede formar una cultura alrededor de temas controvertidos, pero no tan ajenos a la existencia humana como el Big-Data, la inteligencia artificial y el aprendizaje de máquina, entre otros.

Si bien, es normal visualizar que el estadístico es el profesional que la sociedad más vincula con la idea del analista de datos, la Sociedad del dato plantea nuevas demandas que no pueden ser cubiertas únicamente con conocimientos estadísticos; al menos en el contexto colombiano, se están gestando nuevos tipos de analistas de datos como los científicos de datos o los ingenieros en ciencias de datos. Los PECD, en mayor o menor medida, han intentado plantear ofertas que satisfagan estas demandas, con la promesa de preparar profesionales del dato a enfrentar los retos de una sociedad data-céntrica.

Con el fin de definir los diferentes tipos de profesional que conciben los PECD, se puede detectar que el fin último del trabajo con datos parece ser la toma de decisiones, privilegiada por la solución de problemas y el trabajo con personas de otras disciplinas y saberes. La labor del estadístico está más orientada a recopilar, analizar e interpretar datos, aun cuando su tamaño sea importante. Entre tanto, el científico de datos busca darle un valor estratégico a la información contenida en los datos; generar conocimiento y significado a partir de ellos.

La formación de estadísticos, científicos de datos e ingenieros en Ciencia de Datos, al menos desde el punto de vista colombiano, depende de factores como la vigencia del programa o la facultad oferente, lo que dificulta una caracterización del tipo de profesional que se busca formar. Sin pretender que todos los programas de PECD formen profesionales con perfiles idénticos, toda vez que cada universidad brinda más importancia a ciertos valores, creencias o bases características de sus profesionales, y que cobra cada vez más sentido que las facultades reúnan esfuerzos por sacar adelante propuestas de programas de

pregrado, tal vez haga falta un marco común que genere lineamientos en búsqueda de configurar patrones distintivos del profesional del dato.

Desde este trabajo se intentó dar con una conceptualización del análisis de datos, toda vez que esta es una competencia transversal a cualquier persona que se desempeñe en la Sociedad del Dato. Los resultados y discusiones derivadas del análisis que se ha llevado a cabo aquí pueden inspirar investigaciones futuras con relación a la formación de profesionales a nivel general, sin limitarse solamente a los perfiles abordados en este trabajo. Con gran nivel de confianza puede advertirse que la mayoría de los pregrados poseen asignaturas de Estadística en sus planes de estudio, y este trabajo quizá sea el punto de partida para reflexionar acerca de cómo llevar a los profesionales colombianos en formación, a tener competencias en análisis de datos.

Hasta cierto punto, puede admitirse que el rol de analista de datos es la opción más identificable de lo que se deduce de los PECD, visión que resulta contradictoria cuando al verificar que el análisis de datos no ocupa un lugar central en sus propuestas. Lo anterior presenta un panorama preocupante ya que, el lugar donde deberían estar concentrados los esfuerzos por parametrizar el análisis de datos a los pregrados colombianos en general, no tienen una visión unánime ni clara con relación a esta competencia.

En la Sociedad del Dato, aún las carreras que ni se imaginaban usar datos o números en su ejercicio profesional, deben enfrentarse a datos que representan preguntas propias del ejercicio profesional; incluso los más pobres o marginados requieren saber qué hacer ante la información que se deduce de los datos que se generan a su alrededor. Por lo tanto, es deber de las universidades que ofertan

PECD, generar lineamientos que conviertan al análisis de datos en una competencia transversal a todos los niveles educativos, y en particular a todos los profesionales, cuyo contexto de desempeño está permeado por una visión data-céntrica del mundo que les rodea.

Finalmente, quedan todavía bastantes interrogantes por hacer. Es necesario revisar cuál es el valor agregado que están brindando los posgrados en Estadística o Ciencia de Datos en Colombia. Según SNIES, a septiembre de 2022 hay 35 programas de formación posgradual en temas de Estadística, repartidos en 16 especializaciones, 16 maestrías y 3 doctorados. De la misma manera, hay 6 posgrados en Ciencia de Datos, distribuidos en una especialización y cinco maestrías. ¿Cuál es el ámbito o contexto de formación que proyectan con relación a la formación posgradual que están ofreciendo?, ¿En qué medida estas propuestas cualifican el análisis de datos en los posgraduados?

Con toda seguridad, ante el repentino crecimiento de ofertas de pregrados en Ciencia de Datos, habrá que revisar cuál es el papel de los pregrados en Estadística y de esta disciplina en el aparato productivo nacional. Acaso, ¿la Estadística ha quedado relegada para siempre como un campo subsidiario a la Ciencia de Datos?, ¿constituye la Ciencia de Datos un campo de conocimiento con mayor margen de crecimiento o será solamente una cuestión coyuntural que se superará cuando se abra la puerta de una nueva era –por ejemplo, la revolución 5.0 o algo posterior --?, ¿la Estadística ya no tiene respuestas a los problemas más inquietantes de la realidad nacional o mundial?, ¿los estadísticos ya no tienen más remedio que reinventarse como científicos de datos para seguir siendo vigentes en el mercado?.

En búsqueda de garantizar habilidades en análisis de datos para todos los profesionales y en función de los retos que la Sociedad del Dato planteará a la humanidad, y que la gran mayoría no serán reconocidos ni como estadísticos ni como científicos de datos, ¿cómo garantizar, desde los planes de estudio y propuestas curriculares en los pregrados colombianos, que se alcancen estándares suficientes para conseguir una sociedad hábil en el análisis de datos? Dado que, el poder de los datos es algo que todavía la sociedad está tratando de cuantificar y que sus abusos suponen riesgos para la humanidad, ¿de qué manera se puede promover la formación de ciudadanos data-alfabetos?

Referencias bibliográficas

ACOFI. (s. f.). Marco conceptual Matemáticas. Examen de Ciencias Básicas.

Recuperado de <http://www.acofi.edu.co/wp-content/uploads/2018/05/Marco-Conceptual-Matem%C3%A1ticas.pdf>

ACOFI. (2022). Resultados generales Examen de Ciencias Básicas 2021.

Recuperado de <http://www.acofi.edu.co/wp-content/uploads/2022/03/Resultados-Generales-EXIM-2021.pdf>

Alfaro, J., Meseguer, M., Mondéjar, J., & Vargas, M. (2011). Los inicios de la aritmética política en la Inglaterra del siglo XVII. *En: Historia de la probabilidad y la Estadística (V)*. Santiago de Compostela, pp.1-9.

Alfonseca, M. (2000). La máquina de Turing. *En: Las matemáticas del siglo XX una mirada en 101 artículos*. Universidad de la Laguna, 165-168.

Álvarez-Madrigal, M. y López Ríos, O. (2015). Laboratorio para el desarrollo de pensamiento estadístico. Ponencia presentada en el II Congreso internacional de innovación educativa. Monterrey, Tecnológico de Monterrey.

Alzas, T., Casa García, L.M. (2017). La evolución del concepto de triangulación en la investigación social. *Revista Pesquisa Qualitativa*, 5, 395-418.

Anderson, R.D., Sweeney, D.S., Williams, T. A., Camm, J.D., Cochran, J.J. (2019). *Estadística para negocios y economía*. Ciudad de México, México. Cengage Learning.

Andrade, L.; Fernández, F. y Álvarez, I. (2017). Panorama de la investigación en Educación Estadística desde tesis doctorales 2000- 2014. *Revista de la Facultad de Ciencia y Tecnología-Tecné, Episteme y Didaxis, ted*, 41, 87-107.

ARTIST (2006). *Assesment Resource Tools for Improving Statistical Thinking*. Minneapolis, EU, Recuperado el 27 de abril de 2018 de <https://apps3.cehd.umn.edu/artist/about.html>

American Statistical Association (2015). ASA Statement on the Role of Statistics in Data Science. *Amstat News*, 460(9). Recuperado de: <http://magazine.amstat.org/blog/2015/10/01/asa-statement-on-the-role-of-statistics-in-data-science/>

ASA (2015). ASA Statement on the Role of Statistics in Data Science. Recuperado de <https://magazine.amstat.org/blog/2015/10/01/asa-statement-on-the-role-of-statistics-in-data-science/>

ASQ Statistics Division (1996). *Statistical Thinking*, Publicación especial disponible en *American Society for Quality*, Quality Information Center, Milwaukee, WI.

Barahona, A. (2005). Galton y el surgimiento de la genética humana. *Ludus Vitalis*, 23, 151-162.

Barreto-Villanueva, A. (2012). El progreso de la Estadística y su utilidad en la evaluación del desarrollo. *Papeles de Población*, 73, 1-31.

Batanero, C. (2002). Los retos de la cultura Estadística. Presentado en Jornadas Interamericanas de Enseñanza de la Estadística, Buenos Aires.

Batanero, C. (2013). Sentido estadístico: Componentes y desarrollo. Presentado en I Jornadas Virtuales de Didáctica de la Estadística, la Probabilidad y la Combinatoria.

Batanero, C., Díaz, C., Contreras, J., & Roa, R. (2013). El sentido estadístico y su desarrollo. *Números*, 83, 7-18.

BBC (2020). Máquina Enigma: el hallazgo del mítico codificador de mensajes secretos nazi en el fondo del mar. Disponible en <https://www.bbc.com/mundo/noticias-55238523>.

BBVA Openmind (2017). Los fallos que derrotaron a Enigma. Disponible en <https://www.bbvaopenmind.com/tecnologia/innovacion/los-fallos-humanos-que-derrotaron-a-enigma/>

Behar, G., Grima, P. (2004). La Estadística en la educación superior ¿Formamos Pensamiento Estadístico? *Ingeniería y Competitividad*. 5, 84-90.

Bertomeu, J.R. (16 de septiembre de 2020). Investigación y ciencia [Mensaje en un blog]. Recuperado de <https://www.investigacionyciencia.es/blogs/ciencia-y-sociedad/108/posts/william-whewell-y-la-ciencia-como-profesin-liberal-18883>.

Blanco, A. (2018). Directrices y recursos para la innovación en la enseñanza de la Estadística en la universidad: una revisión documental. *Revista de Docencia Universitaria*. 16(1), 251-267.

Bowler, P., Morus, I. (2006). Panorama General de la Ciencia Moderna. Barcelona: Crítica.

- Brown, E.N., Kass, R.E. (2009). What is statistics? *American Statistician*. 62(2), 105-110.
- Budgett, S. y Pfannkuch, M. (2007). Assessing Students' Statistical Literacy. *IASE/ISI Satellite*. Recuperado de https://iase-web.org/documents/papers/sat2007/Budgett_Pfannkuch.pdf
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education - A joint ICMI/IASE study* (pp. 57-69). Dordrecht: Springer.
- Cabañes, M. E., Rubio, R.M. (2013). El sujeto desde la neurociencia y la inteligencia artificial. *Revista estudios de juventud*. 103, 9-19.
- Campos, C.R (2016). La Educación Estadística y la educación crítica. *2° Encuentro Colombiano de Educación Estocástica ECEE*. Bogotá, Colombia.
- Camargo-Vega, J., Camargo-Ortega, J., Joyanes-Aguilar, L. (2015). Conociendo Big Data. *Revista Facultad de Ingeniería*. 38, 63-77.
- Cañón, V., Clavijo, A., Godoy, L., Letouzé, E., Pestre, G., Ricard, J. (2017). Definición de la estrategia de Big-Data para el estado colombiano y para el desarrollo de la industria del Big-Data en Colombia. Disponible en http://datapopalliance.org/wp-content/uploads/2018/09/Documento1_VersionFinal_DNP.pdf

Castro, J. A. (2014). Eugenesia, Genética y Bioética: conexiones históricas y vínculos actuales. *Revista de Bioética y Derecho*, (30), 66-76.

Censo (2019). *EcuRed*. Consultado en enero 5 2021 en <https://www.ecured.cu/index.php?title=Censo&oldid=3482578>.

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*. 10, [3].

Ciampagna, J.M. (2015). Una breve historia sobre Big Data que todo el mundo debería leer. Recuperado de <https://elprofejose.com/2015/03/04/una-breve-historia-sobre-big-data-que-todo-el-mundo-deberia-leer/#:~:text=Nikola%20Tesla%20predice%20que%20en,que%20cabr%C3%A1%20en%20el%20bolsillo>.

Cope, B., Kalantzis, M. (2016). Big Data Comes to School: Implications for Learning, Assessment, and Research. *AERA Open*, 2(2). <https://doi.org/10.1177/2332858416641907>

Corporación Universitaria Iberoamericana. (s.f.). Ingeniería en Ciencia de Datos. Disponible en https://www.iberro.edu.co/wp-content/uploads/2021/10/BROCHURE-INGENIERIA-EN-CIENCIA-DE-DATOS_VIRTUAL.pdf

Cox, M. & Ellsworth, D. (1997). Application-Controlled Demand Paging for Out-of-Core Visualization. Recuperado de <https://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>

Data Science Association (s.f.). Data Science Code of Professional Conduct.

Disponible en <http://www.datascienceassn.org/code-of-conduct.html>

De la Espriella, R., Gómez, C. (2020). Teoría Fundamentada. *Revista Colombiana de Psiquiatría*, 49(2). 127-133.

Del Pino, G., & Estrella, S. (2012). Educación Estadística: relaciones con la matemática. *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, 49(1), 53-64.

De Veaux R.D, Velleman P.F, Bock E.D. (2006). Intro Stats. Boston: Pearson/Addison Wesley. Segunda Edición.

DelMas, R. (2002). Statistical Literacy, Reasoning and Learning: A commentary. *Journal of Statistical Education*. 10, [3].

Denzin, K., Lincoln, Y. (2018). *The SAGE Handbook of Qualitative Research*. (5° ed.). SAGE Publications.

Donoho, D. (2015). 50 years of Data Science. *Journal of Computation and Graphical Statistics*, 745-766.

Escobar, M., Mercado, M. (2019). Big data: un análisis documental de su uso y aplicación en el contexto de la era digital. *Revista La Propiedad Inmaterial*, (28), 273–293. <https://doi.org/10.18601/16571959.n28.10>

Escuela Colombiana de Ingeniería Julio Garavito. (2022). Ingeniería Estadística. Disponible en <https://www.escuelaing.edu.co/es/programas/ingenieria-estadistica/>

- Estrella, S. (2017). Enseñar Estadística para alfabetizar Estadísticamente y desarrollar el razonamiento estadístico. En: Salcedo, A. (Comp.). *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI*, (173- 194). Caracas: Centro de Investigaciones Educativas, Escuela de Educación. Universidad Central de Venezuela.
- Fabris, A., Bartolommei, S., Datteri, E. (2008). Ética y robótica. *Eikasia. Revista de Filosofía*. 19(III), 257-267.
- Fienberg, S.E. (2014). What is Statistics? *Annual Review of Statistics and Its Application*, 1, 1-9. <https://doi.org/10.1146/annurev-statistics-022513-115703>
- Ferreiro, O., & Fernández, P. (1988). La Estadística, una ciencia en la controversia. *Revista Universitaria* (25).
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. (2005). Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework. Alexandria, VA: *American Statistical Association*.
- Franklin, C., Kader, G., Baragliotti, A., Scheaffer, R., Case, C., Spangler, D. (2016). *Statistical Education of Teachers*. Recuperado de <http://www.amstat.org/asa/files/pdfs/EDU-SET.pdf>
- Freedman, D., Pisani, R., Purves, R. (1978). *Statistics*. New York: Norton
- Fresquet, J.L. (2007). Pierre Charles Alexander Louis (1787-1872). Disponible en <https://www.historiadelamedicina.org/louis.html>

Fundación Telefónica (2018). Comunicación codificada. Disponible en <https://espacio.fundaciontelefonica.com/evento/maquina-enigma-comunicacion-codificada/>

Fundación Universitaria Compensar. (s.f.). Ciencia de Datos- Bogotá. Disponible en <https://ucompensar.edu.co/programas/ciencia-de-datos/>

Fundación Universitaria los Libertadores. (2022). Estadística. Disponible en <https://www.ulibertadores.edu.co/facultad-ingenieria/estadistica/>

GAISE College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016*. Recuperado el 11 de marzo de 2018, de <http://www.amstat.org/education/gaise>

Gal, I. (2002). Adult's statistical literacy: Meaning, components, responsibilities. *International Statistical Review* 70(1), 1-25.

Gallardo, S. (2020). Nuevas tecnologías, desde una visión antropológica. *Sistemas*. 154, 24-28.

Galindo, C. (2018). [Yuval Noah Harari, el autor de 'Sapiens': "La tecnología permitirá 'hackear' a seres humanos"]. *El País* (Madrid, España). 29 agosto 2018. https://elpais.com/elpais/2018/08/20/eps/1534781175_639404.html

García, P., Barrero, J. (2019). Actitudes hacia la Estadística de los estudiantes de psicología, ingeniería y economía de la Universidad Católica de Colombia del semestre 2017-I. En Universidad Católica de Colombia (Ed.), *Perspectivas de la enseñanza de la Estadística a nivel universitario* (pp 51-71). Editorial Universidad Católica de Colombia.

Garfield, J. (1999). Thinking about Statistical Reasoning, Thinking, and Literacy. Presentado en The First Annual Roundtable on Statistical Thinking, Reasoning, and Literacy (STRL-1).

Garfield, J. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*. 10(3). Recuperado de www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J, & Ben-Zvi, D. (2008). Research on teaching and learning statistics. En J. B. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and practice* (pp. 21–43). New York, NY: Springer.

González, M. I. (2019). Caracterización de los docentes de Estadística del Departamento de Ciencias Básicas en las facultades de ingeniería y economía de la Universidad Católica de Colombia. En Universidad Católica de Colombia (Ed.), *Perspectivas de la enseñanza de la Estadística a nivel universitario* (pp 35-48). Editorial Universidad Católica de Colombia.

González-Gómez, D. (2014). Constitución de la identidad del profesor que enseña Estadística (Tesis Doctoral). Universidad de Antioquia. Medellín.

Gould, R. (2010). Statistics and the Modern Student. *International Statistics Review*, 78(2), 297-315.

Groth, R. (2015). Working at the boundaries of mathematics education and statistics education communities of practice. *Journal for Research in Mathematics Education*, 46(1), 4–16.

- Guerra, S.M.G., Macedo, G.C., Carvalho Neto, T. M. (2006). Revolución científica, tecnológica y energética: la influencia sobre el pensamiento económico de los siglos XVI Y XVII. *Revista Galega de Economía*, 15(2),1-16. [fecha de Consulta 13 de enero de 2021]. ISSN: 1132-2799. Disponible en: <https://www.redalyc.org/articulo.oa?id=391/39115203>
- Han, B-C. (2014). *Psicopolítica*. Barcelona, España. Herder Editorial S.L.
- Harari, Y. N. (2014). *De animales a dioses. Breve Historia de la Humanidad*. Barcelona: Debate.
- Harari, Y. N. (2015). *Homo Deus*. Bogotá: Penguin Random House.
- Harari, Y. N. (2017). *21 lecciones para el siglo XXI*.
- Haraway, D. J. (1995). Ciencia, cyborgs y mujeres. La invención de la naturaleza, Cátedra, Madrid.
- Harnad, S., (1992). The Turing Test is Not a Trick: Turing Indistinguishability Is a Scientific Criterion. *Boletín ACM SIGART*, 3(4), 9-10.
- Hicks, S. C., Irizarry, R. A. (2018). A Guide to Teaching Data Science. *The American statistician*, 72(4), 382–391. <https://doi.org/10.1080/00031305.2017.1356747>
- Hoerl, R. W., Snee, R. (2010). Statistical Thinking and Methods in Quality Improvement: A Look to the Future, *Quality Engineering*, 22: 3, 119 -129.
- Hoerl, R., Snee, R., & De Veaux, R. (2014). Applying statistical thinking to Big Data Problems. *WIREs Comput Stat*, 6, 222-232.

Holik, F. (2016). Teoría de la información de Claude E. Shannon. En Diccionario Interdisciplinar Austral, editado por Claudia E. Vanney, Ignacio Silva y Juan F. Franck.

URL=http://dia.austral.edu.ar/Teoría_de_la_información_de_Claude_E._Shannon

ICFES. (2017a). *Guía de orientación módulos de competencias genéricas SABER-PRO-2017*. Bogotá. Recuperado el 28 de marzo de 2018, de <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/guias-de-orientacion>

ICFES. (2017b). *Guía de orientación Saber Pro-Competencias Específicas Módulos de Pensamiento Científico*. Recuperado el 28 de marzo de 2018, de <http://www.icfes.gov.co/instituciones-educativas-y-secretarias/saber-pro/guias-de-orientacion>

ICFES. (2020a). *Informe de resultados módulos genéricos Saber Pro-2019*. Bogotá. Recuperado el 22 de abril de 2021 en <https://www.icfes.gov.co/web/guest/resultados-saber-pro>

ICFES. (2020b). *Informe de resultados módulos específicos Saber Pro-2019*. Bogotá. Recuperado el 22 de abril de 2021 en <https://www.icfes.gov.co/web/guest/resultados-saber-pro>

ICFES. (2021). *Informe nacional de resultados del examen Saber Pro-2020*. Bogotá. Recuperado de https://www2.icfes.gov.co/web/guest/resultados_del_examen_saber_pro.

- INEI. (2006). Glosario básico de términos estadísticos. Lima. Recuperado de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0900/Libro.pdf
- Ito, P.K. (2000). Respuesta a Investigación en Educación Estadística: Algunas Cuestiones Prioritarias. *Statistical Education Research Newsletter*. 1(2).
- Iversen, G. R, Gergen M. (1997). *Statistics: The Conceptual Approach*. New York: Springer.
- Joyanes-Aguilar, L. (2013). *Big Data Análisis de grandes volúmenes de datos en organizaciones*. México. Alfaomega Grupo Editor.
- Kemp, S. (2022). Digital 2022: Global Overview Report. Recuperado de <https://datareportal.com/reports/digital-2022-global-overview-report>
- Kolb, S. (2012). Grounded theory and the constant comparative method: Valid research strategies for educators. *Journal of Emerging Trends in Educational Research and Policy Studies*. 3. 83-86.
- Lateralía (2016). *Del humanismo al dataísmo*. Recuperado de <https://lateralía.es/del-humanismo-al-dataismo/>
- Llanos, D. (2010). *Fundamentos de lógica y programación en C*. Paraninfo. Madrid, España.
- Lohr, S. (2015). *Data-ism*. Nueva York: HarperCollins Publisher.
- Lupton, D. (2016). *The Quantified Self*. Cambridge, Reino Unido. Polity Press.

- MacKay, J. (2016). Discussion: Acquiring Statistical Literacy and Thinking. *International Statistical Review*, 84(2), 189-194.
- Makar, K., Confrey, J. (2005). Using distributions as statistical evidence in well-structured and ill-structured problems. In Reasoning About Distribution: A Collection of Current Research Studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-4), Ed. K. Makar, University of Auckland, New Zealand, 2-7 July. Brisbane: University of Queensland.
- Martínez, M. (2004). Ciencia y Arte en la Metodología Cualitativa. México. Editorial Trillas.
- Martínez, G. (2014). Una propuesta didáctica para fortalecer las competencias de lectura y construcción de tablas y gráficos estadísticos. Universidad Sergio Arboleda, Bogotá.
- Martínez, G., Soto, W. (2019). Pedagogía del dato: una perspectiva desde la enseñanza de la Estadística en la sociedad del dato. *Análisis*. 94. 141-158.
- Mateo, A. (2016). La pervivencia clásica del censo de los ciudadanos romanos. *Persona y Derecho*, 75(2), 57-97.
- Marradi, A., Archenti, N., Piovani, J.I. (2007). Metodología de las ciencias sociales. Buenos Aires, Argentina. Emecé Editores.
- Mayer-Schönberger, V., Cukier K. (2013). Big Data: A revolution that will Transform How We Live, Work, and Think. New York, Estados Unidos. Houghton Mifflin

- Medina Lloret, C. (2021). El estado del arte de la Ciencia de Datos en el idioma español y su aplicación en el campo de la Inteligencia Artificial. En TM. <https://doi.org/10.21428/39829d0b.f5257ea7Harcourt>.
- Mosteller, F., Rourke, R.E.K., Thomas, G.B Jr., (1961). *Probability with Statistical Applications*. Reading, MA: AddisonWesley
- Moore, D. S. (1997). Mathematics, Statistics and Teaching. *The American Mathematical Monthly*, 104: 801-823.
- Moore, D. S. (1998), Shaping Statistics for Success in the 21st Century: A Panel Discussion, Kansas State University Technical Report II-98-1.
- Morabia A. (2006). Pierre-Charles-Alexandre Louis and the evaluation of bloodletting. *Journal of the Royal Society of Medicine*, 99(3), 158–160. <https://doi.org/10.1258/jrsm.99.3.158>.
- Nikiforidou, Z., Lekka, A., Pange, J. (2010). Statistical literacy at university level: the current trends. *Procedia Social and Behavioral Sciences*, 9, 795-799.
- OCDE. (2015). *Guía para una revolución de los datos dirigida por los países Paris 21*. Paris.
- O'Neill, C. (2016). *Armas de destrucción matemática*. Capitan Swing Libros.
- ONU. (2015). A new global partnership: eradicate poverty and transform economies through sustainable development. Disponible en [13 HLP UN-Report extracts.pdf](#)

- Oppenheimer, A. (2018). ¡Sálvese quien pueda! Ciudad de México. Penguin Random House.
- Pantula, S. (2011). Statistics: A key to innovation in a Data-Centric World! *Journal of the American Statistical Association*, 493, 1-5.
- Páramo, D. (2015). La teoría fundamentada (Grounded Theory), metodología cualitativa de investigación científica. *Pensamiento & Gestión*, (39), vii-xiii.
- Paredes-Moreno, A. (2015). Big Data: Estado de la cuestión. *International Journal of Information Systems and Software Engineering for Big Companies (IJISEBC)*, Vol. 2, Num. 1, pp. 38-59.
- Petrelli, M. (2018). Cómo gerenciar el cambio tecnológico en la nueva revolución industrial. *AADECA Revista*, 10, 12-14.
- Pérez, L., Aparicio A., Bazán, J., y Abdounur, O. (2015). Actitudes hacia la Estadística de estudiantes universitarios de Colombia. *Educación Matemática*, 27(3), 111-149.
- Pinto, J., Tauber, L., Zapata-Cardona, L., Albert, A., Makofazi, J. (2017). Alfabetización Estadística en educación superior. *Acta Latinoamericana de Matemática Educativa*, 227-235.
- Piovani, J. I. (2007). Los orígenes de la Estadística: de investigación sociopolítica empírica a conjunto de técnicas para el análisis de datos. *Revista de Ciencia Política y Relaciones Internacionales*, 1(1), 25-44.

Politécnico Gran Colombiano. (s. f.). Ingeniería en Ciencia de Datos Bogotá.

Disponible en <https://www.poli.edu.co/profesional/ingenieria-en-ciencia-de-datos-bogota>

Pontificia Universidad Bolivariana. (s. f.). Ingeniería en Ciencia de Datos. Disponible

en <https://www.upb.edu.co/es/pregrados/ingenieria-ciencia-datos-medellin>

Pontificia Universidad Javeriana. (s. f.). Ciencia de Datos. Disponible en

<https://www.javeriana.edu.co/carrera-ciencia-de-datos>

Preciado-Vidal, G. (2019). El big-data y la huella digital: la importancia de los datos y cómo son utilizados por las empresas. Universidad Pontificia Comillas. Madrid, España.

Proctor, R. (1992). Nazi Doctors, Racial Medicine and Human Experimentation. En

G. Annas and A. Grodin. The Nazi doctors and the Nuremberg code (pp. 17-31). New York, USA: Oxford University Press.

Quintela del Río, A. (2019). Estadística Básica Edulcorada. Recuperado de

<https://bookdown.org/aquintela/EBE/>

Ramos, L. (2019). La Educación Estadística en el nivel universitario: retos y

oportunidades. Revista Digital de Investigación en Docencia Universitaria, 13(2), 67- 82.

Riascos, Y. (2014). El pensamiento estadístico asociado a las medidas de tendencia

central: un estudio psicogenético sobre la media aritmética, la mediana y la moda (Tesis Doctoral). Universidad del Valle, Cali.

- Riascos, Y. (2016). Razonamiento estadístico y otros conceptos relacionados. 2° *Encuentro Colombiano de Educación Estocástica ECEE*. Bogotá, Colombia.
- Roca, T. Letouzé, E. (2016). Open algorithms: A new paradigm for using private data for social good. Disponible en [Open algorithms: A new paradigm for using private data for social good | Devex](#)
- Rocha, P. (2012). La Educación Estadística en la formación de ingenieros. *Educación Científica y Tecnológica*, 17, 33-45.
- Romero, M. (2018). En Estadística no importa el tamaño hasta que deja de hacerlo: small & big-data. *Pharmatech*. 34. 34-50.
- Ross, S.M. (1996). *Introductory Statistics*. New York: McGraw-Hill.
- Ruiz, G.J. (2006). Historia de las Estadísticas de salud. *Gaceta Médica Boliviana*, 29(2), 72-77.
- Ruiza, M., Fernández, T. y Tamaro, E. (2004). Biografía de Herman Hollerith. En *Biografías y Vidas. La enciclopedia biográfica en línea*. Barcelona (España). Recuperado de <https://www.biografiasyvidas.com/biografia/h/hollerith.htm> el 10 de marzo de 2021.
- Rumsey, D. J. (2002) Statistical Literacy as a Goal for Introductory Statistics Courses" *Journal of Statistics Education* [Online], 10(3). www.amstat.org/publications/jse/v10n3/rumsey2.html
- Saade, M.G. (2004). ¿Quiénes deben procrear? Los médicos eugenistas bajo el signo social (México, 1931-1940). *Cuicuilco*, 31(11), 0.

- Sánchez, J. (2013). Criptología Nazi. Los códigos secretos de Hitler. *Pensamiento Matemático*. 3(1). 59-119.
- Sandín, M. (2003). *Investigación cualitativa en educación. Fundamentos y Tradiciones*. México: Mc Graw Hill.
- Schild, M. (2010). Assessing Statistical Literacy: Take CARE. En P. Bidgood, N. Hunt, & F. Jolliffe (eds.) *Assessment Methods in Statistical Education: An International Perspective*, 133-152. John Wiley & Sons Inc: Chichester, Great Britain.
- Schwab, K. (2016). The Fourth Industrial Revolution: what it means, how to respond. Recuperado el 15 de febrero de 2021 de <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- Schweidel, D.A. (2015). *Profiting from the data economy*. New Jersey. Pearson Education.
- Science (2011). Dealing with data. Special Online Collection, 311. Recuperado de: <http://www.sciencemag.org/site/special/data/>
- Serradó Bayes, A. (2013). El Proyecto Internacional de Alfabetización Estadística. *Números*, 83, 19-33.
- Sibilia, P. (2006). *El hombre postorgánico*. Buenos Aires: Fondo de Cultura Económica.

- Sierra-Bonilla, Y., Rodríguez, A. (2019). Caracterización de los docentes de Estadística del departamento de ciencias básicas en las facultades de ingeniería y economía de la Universidad Católica de Colombia. En Universidad Católica de Colombia (Ed.), *Perspectivas de la enseñanza de la Estadística a nivel universitario* (pp 35-48). Editorial Universidad Católica de Colombia.
- Siles, I. (2007). Cibernética y sociedad de la información: El retorno de un sueño eterno. *Signo y pensamiento*, 25(50), 84-99.
- Snee (1990). Statistical Thinking and Its Contribution to Total Quality. *The American Statistician*, 44, 116-121.
- Snell, L. (1999). Using Chance media to Promote Statistical Literacy. Presentado en the 1999 Joint Statistical Meetings, Dallas, TX.
- Soto, W. (2013). Cibernética y Pedagogía. Una mirada desde la física y la matemática hacia la pedagogía del algoritmo. Universidad de San Buenaventura, Bogotá.
- Stohl, J. (2005). Probability in teacher education and development. In Exploring Probability in School: Challenges for Teaching and Learning, Ed. G.A. Jones, pp. 345-366. New York: Springer.
- Strauss, A. (2004) Anselm Strauss en conversación con Heiner Legewie y Barbara Schervier-Legewie. Forum Qualitative Social Research,(on line journal) 5 (3) Art. 22. Disponible en: <http://www.qualitative-research.net/fqgstexte/3-04/04-3-22b-s.htm>

Strauss, A., Corbin, J. (1998). *Basics of qualitative research. Techniques and procedures for developing grounded theory*. Los Ángeles. SAGE Publications.

Tabares, L.F., Hernández, J.F. (2014). *Big Data Analytics: oportunidades, retos y tendencias*. Cali. Universidad de San Buenaventura.

Tukey, J. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*. 33(1), 1-67.

Universidad de Antioquia. (2020). Pregrado en Estadística. Disponible en <https://www.udea.edu.co/wps/portal/udea/web/inicio/unidades-academicas/ciencias-exactas-naturales/estudiar-facultad/pregrados/estadistica>

Universidad del Norte. (s.f). Pregrado en Ciencia de Datos Disponible en <https://www.uninorte.edu.co/en/web/ciencia-de-datos>

Universidad ECCI. (s.f.). Ciclo único. Programa en Estadística. Disponible en <https://www.ecci.edu.co/programas/estadistica/>

Universidad el Bosque. (s.f.). Estadística. Disponible en <https://www.unbosque.edu.co/ciencias/carrera/estadistica>

Universidad Externado de Colombia. (s.f). Pregrado en Ciencia de Datos. Disponible en <https://www.uexternado.edu.co/programa/departamento-de-matematicas/ciencia-de-datos/>

Universidad INCCA de Colombia. (2022). Programa profesional en Estadística. Disponible en <https://unincca.edu.co/programa/programa-profesional-de-estadistica/>

Universidad Nacional de Colombia. (2015). Proyecto Educativo del programa de Estadística. Disponible en http://www.pregrado.unal.edu.co/docs/pep/pep_2_7.pdf

Universidad Santo Tomás. (2017). Facultad de Estadística. Recuperado de <https://facultadestadistica.usta.edu.co/>

Universidad del Valle. (2017). Proyecto Educativo del programa de Estadística. Disponible en <http://escuelaestadistica.univalle.edu.co/pregrado-en-estadistica>.

UNESCO. (2011). *Alfabetización mediática e informacional. Currículum para profesores*. Paris.

Universidad Rómulo Gallegos (2013). Comprensión y uso de la Estadística (glosario). Disponible en <https://web.cortland.edu/flteach/stats/glos-sp.html>

Utts, J., Horton, N. I. (2016). The role of Statistics in Data Science, and vice versa. Presentation at the Roundtable on Data Science Post-Secondary Education. National Academics of Science. Washington D.C. Recuperado de: http://sites.nationalacademies.org/DEPS/BMSA/DEPS_175092

Vasquez, A. (2020). Ciencia de Datos para Gente Sociable.

- Valero, V. (2018). Big-Data: cómo afecta a la privacidad de los ciudadanos. Universitat Oberta de Catalunya. España.
- Vere-Jones, D. (1997). The coming of age of statistical education. *International Statistics Review*, 63. 3-23.
- Villela, F.C., Linares, J, E. (2011). Eugenesia: un análisis histórico y una posible respuesta. *Acta Bioethica*, 17(2). 189-197.
- Virseda, F., Román, J. (s.f.). Minería de Datos y aplicaciones. Universidad Carlos III. Madrid
- Wallman, K. K. (1993). Enhancing statistical literacy: enriching our society. *Journal of the American Statistical Association*, 88, 1-8.
- Watson, J. (1997). Assessing statistical thinking using the media. En I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education*. (pp. 107-121). Amsterdam, The Netherlands: IOS Press.
- Watson, J. M. (2001). Profiling teachers' competence and confidence to teach mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, 4, 305-337.
- Watson, J. y Callingham, R. (2003). Statistical Literacy: a complex hierarchical context. *Statistical Education Research Journal*, 2(2), 3-46.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, N.J.: Erlbaum Associates.

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics*. 96, 33-47.

Wiener, N. (1998a). *Cibernética: o el control y comunicación en animales y máquinas*. Barcelona: Tusquets editores.

Wiener, N. (1998b). *Dios y Golem*, S.A. México: Siglo XXI editores.

Wild, C., Triggs, C., Pfannkuch, M. (1997). Assessment on a budget: using traditional methods imaginatively. En Gal, I., Garfield, J. (eds), *The Assessment Challenge in Statistics Education*, 205-220.

Wild, C., Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67 (3), 221-248.

Wild, C., Pfannkuch, M. (2000). Statistical thinking in empirical enquiry. *Statistical Science*, 15 (2), 132-252.

Wilks, S. (1951). Undergraduate Statistical Education. *Journal of the American Statistical Association*, 46, 1-18.

Wills, J. (3 de mayo de 2012). *Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any* [Tweet].

https://twitter.com/josh_wills/status/198093512149958656?cxt=HHwWgICQgP6_4r8FAAAA

WIRED. (2014). *Data is the New Oil of Digital Economy*. Recuperado el 28 de marzo de 2018, de <https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/>

Wu, C.F. (s.f.). Statistics= Data Science? Recuperado de <http://www2.isye.gatech.edu/~jeffwu/presentations/datascience.pdf>

Yañez, S. (2000). La Estadística, una ciencia del siglo XX, Fisher el genio. *Revista Colombiana de Estadística*, 23(2).

Youyou, W. Kosinski, M., Stillwell, D. (2014). Computer- based personality judgments are more accurate than those made by humans. *PNAS*, 112. 1036-1138.

Zapata, L. (2011). ¿Cómo contribuir a la alfabetización Estadística? *Revista Virtual Universidad Católica del Norte*, 33, 234-247. Obtenido de <http://www.redalyc.org/articulo.oa?id=194218961013>

Zapata, L. (2016). ¿Estamos promoviendo pensamiento estadístico en el aula? 2° *Encuentro Colombiano de Educación Estocástica ECEE*. Bogotá, Colombia.

