

UNA APROXIMACIÓN A LA INVESTIGACIÓN MULTI E INTERDISCIPLINAR: DIVERSOS ENFOQUES TEÓRICOS Y METODOLÓGICOS

2DA JORNADA DE INVESTIGACIÓN
Y ACTUALIZACIÓN INTERNA UVM

**UNA APROXIMACIÓN A LA INVESTIGACIÓN
MULTI E INTERDISCIPLINAR:
DIVERSOS ENFOQUES TEÓRICOS
Y METODOLÓGICOS**

**2DA JORNADA DE INVESTIGACIÓN
Y ACTUALIZACIÓN INTERNA UVM**

**UNA APROXIMACIÓN A LA INVESTIGACIÓN MULTI E INTERDISCIPLINAR:
DIVERSOS ENFOQUES TEÓRICOS Y METODOLÓGICOS.
2º Jornada de investigación y actualización interna UVM**

© Adriana Villa-Murillo
© Mauricio Echiburú F.
© José Rivera-Soto
© Alfredo González S.
© Camilo Henríquez Miranda
© Corina González-Weil
© Paulina Bravo González
© Silvia Soledad López
© Patricia Cabrera-Calavacero
© Neumarino Rodríguez V.
© Matías Montesinos
© Catalina León Martínez
© Danilo Reyes-Lillo
© Universidad Viña del Mar 2022

ediciones@uvm.cl

Director de la colección: José Rivera-Soto

Director de arte: Xavier Adaros Manríquez

Diseño y diagramación: Daniella Morales

Viña del Mar, Chile.

ISBN: 978-956-8135-18-8

Ediciones UVM



PRESENTACIÓN

Con el fiel compromiso de fomentar y actualizar el quehacer investigativo de los estudiantes y académicos de nuestra Universidad, se pone en marcha la 2da Jornada de investigación y actualización interna UVM, esta vez con un objetivo concreto en cuanto a algunos enfoques teóricos y metodológicos de la investigación multi e interdisciplinaria. Tales ideas se presentan en el texto: **Una aproximación a la investigación multi e interdisciplinar: diversos enfoques teóricos y metodológicos**, como el principal aporte a dicho evento.

En efecto, el texto está conformado por tres ejes: en primer lugar se introducen los aspectos más resaltantes de la investigación formativa, desde su definición, alcances y limitaciones, hasta su instalación y aportes con su respectivo programa en la UVM. Para una mejor comprensión de tales aspectos, se presenta un caso concreto de estudio.

La segunda parte da una mirada más disciplinar a la investigación mediante aportes de algunos académicos: temas centrados en la inclusión social, la didáctica de la matemática y la neurociencia, así como tópicos más específicos como la media derivada desde un estudio socio epistemológico y la modelización de la evolución de un disco protoplanetario, los cuales pretenden inducir e invitar al lector a reflexionar sobre la importancia de grupos multi e interdisciplinarios de investigación.

Una tercera y última parte, se concentra en el proceso de actualización académica en nuestro objetivo de alfabetización estadística, iniciado en las jornadas 2021. En esta oportunidad se introduce al lector en la gestión de los datos de investigación mediante consideraciones básicas, seguidamente se aborda el manejo de datos estadísticos desde la base descriptiva (fase irrenunciable en cualquier tipo de investigación) y se introduce el análisis de regresión y correlación lineal a través de sus aspectos metodológicos.

Desde el programa de investigación de la Escuela de Ciencias, la Coordinación de Investigación y Postgrado y el CREA (Centro de recursos para la enseñanza y aprendizaje) nos complace compartir con la comunidad nuestros aportes e ideas en torno a la investigación desde diferentes aristas, como un punto de encuentro y enriquecimiento para el quehacer académico y social.

Dra. Adriana Villa-Murillo

TABLA DE CONTENIDOS

Nota editorial	7
PRIMERA PARTE: Investigación formativa	
Investigación formativa en la Universidad Viña del Mar, Mauricio Echiburú F.	10
Creación e investigación formativa en la Universidad Viña del Mar. Aproximaciones a la instalación, aportes y mecanismos de una noción emergente, José Rivera-Soto y Alfredo González S.	16
Análisis de las historias de clase de profesores que reflexionan para una educación transformadora en ciencias: estructura y cambios en el tiempo, Camilo Henríquez Miranda, Corina González-Weil, Paulina Bravo González.	25
SEGUNDA PARTE: La investigación desde la óptica multi e interdisciplinar	
Paradigmas de la didáctica de las Matemáticas y su aplicación en particular al objeto matemático de optimización, Silvia Soledad López	32
Un acercamiento a la Neuroeducación, Patricia Cabrera-Calavacero	45
La media derivada, Neumarino Rodríguez V.	52
Presión de radiación y sombras en discos protoplanetarios, Matías Montesinos	56
TERCERA PARTE: Manejo de datos y actualización estadística	
Encontrables, accesibles, interoperables y reutilizables: algunas consideraciones para la gestión de nuestros datos de investigación, Danilo Reyes-Lillo	67
Estadística Descriptiva: el primer paso de toda investigación, Catalina León Martínez	76
Usos y abusos en los análisis de regresión y correlación lineal, Adriana Villa-Murillo	97

NOTA EDITORIAL

Para esta segunda versión de las Jornadas de investigación y actualización interna de la UVM, quisiera iniciar esta nota editorial con una pequeña reflexión. A saber, en contexto entendemos a la investigación como toda actividad intelectual dedicada a la búsqueda sistemática, planeada, autocrítica y metódica del conocimiento que se expresa en su comprensión, diseño y generación y que permite preservar y mejorar la calidad de vida del ser humano y su entorno. También es entendida como la abstracción de un creador, inmerso en entornos culturales, sociales y políticos, que materializa una obra, que reconoce los sentidos y construye tejidos sociales.

Por ello y a nuestro entender, la investigación debe traspasar todas las áreas específicas del saber en la búsqueda de respuestas a situaciones humanas cotidianas y trascendentales, a través de métodos y sistemáticas creativos, productos del constante desarrollo y generación de conocimiento. Conocimiento que tradicionalmente ha estado alojado en las instituciones de educación superior, pero que cada vez más instituciones privadas y públicas extrauniversitarias han ido incursionando en la constante búsqueda del conocimiento, el saber y el desarrollo de los entornos culturales y los sentidos que la comunidad les da.

Como toda institución de educación superior, la UVM no está ajena a este quehacer y ha debido desarrollar actividades de investigación en todas sus áreas, expresadas en un conjunto significativo de proyectos de investigación integrados en la comunidad y por cierto, al sistema nacional de ciencia y tecnología.

En la UVM en estos últimos años, la investigación básica y aplicada como también la investigación formativa, se han venido desarrollando de forma sistematizada a través de una política, normativas y focos estratégicos establecidos por la dirección de investigación y postgrado, como también en forma autodidacta y a discreción por académicos que tienen las redes de contacto, inquietud natural o formación para aquello. Estas actividades se han desarrollado en variadas áreas del conocimiento y han generado redes de contacto e investigación, con importante participación de docentes en proyectos con financiamiento externo, privado y publicaciones de divulgación científicas y profesionales.

Así, estas jornadas de investigación y actualización interna de la UVM en su segunda versión plantea “fomentar y actualizar el quehacer investigativo de los estudiantes y académicos, con el objetivo de establecer equipos multidisciplinarios y/o semilleros de investigación”. De forma específica se buscará actualizar y conocer lo que se realiza en investigación formativa, sus principales líneas, aportes y aplicaciones; en líneas disciplinares se planteará la invitación para conformación de equipos interdisciplinarios para futuras investigaciones; se actualizará y buscará la reflexión de la gestión de datos y estadística, como un aporte en el fomento de las buenas prácticas y pensamiento crítico; finalmente se buscará fomentar

el intercambio de ideas y/o experiencias entre académicos y definir líneas de acción que sustenten las próximas jornadas de carácter formativo y de actualización investigativa.

Estas jornadas, como importante hito, seguirán siendo el foco de aglutinación en torno a la discusión de las materias de investigación en todas sus vertientes disciplinares y formativas, e invitarán también a generar la necesaria interdisciplinariedad y semillero para las ideas e inquietudes que muchos académicos y estudiantes noveles aún tienen en investigación.

Dr. Francisco J. González

PRIMERA PARTE

INVESTIGACIÓN FORMATIVA

1

Investigación formativa en la Universidad Viña del Mar

Mauricio Echiburu

*Escuela de Ciencias
mechiburu@uvm.cl*

RESUMEN

La Investigación Formativa (IF) es un concepto relativamente nuevo en el contexto universitario, muchas veces mal entendido o con definiciones muy amplias que generan desconocimiento de parte de académicos y estudiantes. En este trabajo se pretende explicar los alcances y resultados que se han logrado con el programa de IF de la Universidad Viña del Mar (UVM), definiendo qué se entiende por IF, con qué recursos previos se cuenta, cuáles son los objetivos y líneas de investigación del programa, detallando los hitos que se han logrado en este primer año de funcionamiento.

Palabras clave: Investigación formativa, docencia, investigación educativa, innovación educativa.

Introducción

Es ampliamente aceptado por las comunidades educativas y la sociedad civil en general, que la investigación está íntimamente ligada con la calidad de la formación en Instituciones de Educación Superior (IES), entendiendo investigación y creación por todo proceso que se refiere a la búsqueda y generación de nuevos conocimientos. Para IES como la Universidad Viña del Mar (UVM), que presenta una emergente experiencia en investigación de alto nivel, se hace extremadamente necesario no solo realizar investigación, sino además ser capaces de utilizar dichos productos pertinentemente en docencia.

Más aún, considerando que el Consejo Nacional de Acreditación (CNA) exige como uno de sus criterios de acreditación de carreras y programas de pregrado que las y los docentes desarrollen Investigación Formativa (IF). A nivel país y en términos generales, la IF o es mal entendida por la academia o no es considerada como investigación propiamente tal. Por fortuna, estos antecedentes han ido cambiando en los últimos tiempos, haciendo que la IF comience a ser parte importante de la docencia, contando con prácticas que han sido desarrolladas por las instituciones de educación y por sus docentes en forma efectiva. Lamentablemente, muchas de estas prácticas no quedan registradas, no se realizan seguimientos, o no se miden los alcances de estas. Por ello, el programa de Investigación Formativa de la UVM, iniciado en enero del 2021, pretende fortalecer todos los aspectos de la práctica docente en estas áreas, llevando en un futuro cercano a la formación de un Centro de Investigación en Educación dentro de la UVM que cubra tanto la investigación en educación como en didáctica.

Desarrollo

En la actualidad, la relación entre docencia e investigación, pasa por la precisión del concepto de Investigación Formativa y la Investigación Científica propiamente tal. En la segunda mitad de la década del 90 se comenzó a hablar de IF como:

Aquel tipo de investigación que se hace entre estudiantes y docentes en el proceso de desarrollo del currículo de un programa y que es propio de la dinámica de la relación con el conocimiento que debe existir en todos los procesos académicos tanto en el aprendizaje, por parte de los alumnos, como en la renovación de la práctica pedagógica por parte de los docentes (Restrepo, 2003).

Esta definición, no ha sido todo lo precisa que se hubiese deseado, y genera aún cierto grado de discusión en los marcos interpretativos que contienen y delimitan la IF. Por ello, hay tres acepciones de IF que definen mejor los límites a los que debemos atenernos:

- Investigación exploratoria: se refiere a la investigación que se realiza para buscar las necesidades, problemas, hipótesis y poblaciones relevantes para estructurar o refinar proyectos de investigación. Se trata de “dar forma” a la investigación, no de la formación del estudiante.
- Formación en y para la investigación: a través de actividades, que no necesariamente sean parte de un proyecto concreto de investigación, se familiariza al estudiante con la investigación.
- Investigación acción: investigación realizada para aplicar sus hallazgos sobre la marcha, para afinar y mejorar los programas mientras están siendo desarrollados.

La IF, orientada a la mejora de las teorías y las prácticas de enseñanza por parte del cuerpo docente de la UVM, se fundamenta en reglamentos que se han impulsado en los últimos años como la Política de investigación de la Universidad Viña del Mar (Dirección de Investigación y Postgrado, 2019). Esta política orienta el quehacer investigativo con el propósito de vincular la investigación con programas de pre y postgrado y el desarrollo de estudiantes y académicos, así como también el favorecer y valorar la investigación, especialmente en el área formativa.

En esta línea se han desarrollado proyectos de innovación docente y existe una gran oportunidad de llevar a trabajos de investigación, al guiar seminarios de título en pregrado y proyectos de grado en posgrado, cuando la o el profesor guía cumple una verdadera función de asesor en líneas investigativas. Además, no solo al final de la carrera del estudiante se puede realizar IF, también los ensayos teóricos, en diferentes disciplinas, son un recurso que puede ser utilizado bajo un esquema de trabajo de investigación. Asimismo, se puede asociar a los estudiantes con la investigación que realizan los docentes que investigan, logrando así que estos formulen soluciones a problemas y diseñen proyectos. En la Universidad de Michigan, por ejemplo, vinculan a los estudiantes a proyectos de investigación de los docentes, utilizando para ello fondos de investigación donados por fundaciones o empresas. Los estudiantes asumen un compromiso adicional al de su plan de estudios y deben entregar informes oportunos sobre su trabajo (Neihardt, 1997).

El mejoramiento de la docencia a través de la indagación como investigación logra en la práctica el aprendizaje como construcción del conocimiento por parte de las y los estudiantes, como una consecuencia del mejoramiento de los cursos y programas académicos. Ejemplos específicos de esto tenemos en la implementación de estrategias como el Aprendizaje Basado en Problemas (ABP), Aprendizaje Basado en Proyectos, portafolios, búsqueda y reseña bibliográfica a través de “clubes de revistas” o cuando se supera la evaluación de contenidos y se evalúan procesos superiores de pensamiento, centrándose en habilidades y competencias de orden superior.

Un aspecto que también se debe considerar es la enorme capacidad y competencia de las y los docentes UVM para asesorar, tanto en forma interna como externa, en cuestiones de educación y didáctica. Existe múltiple evidencia sobre cursos y capacitaciones referentes a evaluación, estrategias docentes, TIC`s y otros, que se han realizado para académicos al interior de las Escuelas y Departamentos en la Universidad, así como también, experiencias y proyectos que han llevado estas capacitaciones y asesorías a todos los niveles educativos tanto dentro como fuera del país. Esto nos lleva a pensar, en que este programa y en el corto plazo centro de investigación, lograría entregar asesorías y certificaciones educacionales a todo nivel de enseñanza.

No se puede dejar de nombrar, la importante misión social de la educación. Desde este punto de vista, la IF es clave en el trabajo directo con la comunidad. Por ejemplo, a través de iniciativas como el Aprendizaje Servicio (A+S) se logra asesorar, diagnosticar y dar soluciones a la comunidad por medio de proyectos de investigación desarrollados por docentes y estudiantes. Específicamente, en la Universidad Viña del Mar este rol es apoyado por la Vinculación con el Medio (VcM), que es considerada como una función esencial del que hacer institucional ya que se define como el conjunto de lazos que emergen a través de acciones desarrolladas entre la universidad e instituciones privadas, públicas y de la sociedad civil, en su entorno significativo. El propósito de esta interacción se basa en el enriquecimiento mutuo, colaboración, co-creación, compromiso, evaluación y retroalimentación positiva del quehacer universitario y del territorio (2019, p. 12).

Es así como en la Universidad Viña del Mar la Vinculación con el Medio ha evolucionado desde una nutrida agenda de extensión cultural y universitaria a instancias de vinculación con mayores atributos de bidireccionalidad, siendo la IF una prioridad, incrementado la relevancia de la vinculación con el medio en el proceso formativo de los estudiantes.

El objetivo general del programa de IF en nuestra Universidad es desarrollar la investigación formativa en la Universidad, sea tanto como investigación exploratoria, investigación para la formación y/o en investigación acción.

Los objetivos específicos del Programa de Investigación Formativa son:

- Promover la innovación educativa en los procesos de formación de la universidad.
- Generar acciones que lleven a líneas definidas de investigación formativa y motivar el conocimiento de estas líneas dentro del cuerpo docente.
- Fortalecer tanto dentro como fuera de la UVM las redes de cooperación, sean estas a nivel nacional y/o internacional.
- Incentivar la productividad en IF a través de proyectos, tesis de grado, publicaciones y otros.

Estos objetivos se cumplen a través de las líneas de investigación ha desarrollar en forma conjunta por las diferentes escuelas y departamentos de la universidad. Líneas que se pueden resumir en los siguientes puntos:

- Diseño, implementación y/o seguimiento de innovaciones educativas en formación de pre y postgrado.
- Vinculación con el medio a través de alianzas universidad-comunidad-empresas.
- Formaciones diversas en IES: jornadas vespertinas, continuidades de estudio, multimodalidad, educación continua.
- Evaluación y gestión en educación superior.
- Educación inclusiva en educación superior.
- Problemas de las IES: retención, progresión estudiantil y eficacia académica.
- TICs para la educación superior.
- Educación de adultos.
- Metodologías de enseñanza integradas (STEM + otras)

Las líneas de investigación asociadas a este programa se vinculan con dos grandes dimensiones. Una de ellas se refiere a las políticas y gestión en educación y la otra se centra en didáctica y estrategias docentes.

En una primera etapa el programa se centra en promover e incentivar la IF ya existente y promover las nuevas iniciativas desde la mirada del diseño, implementación y seguimiento tanto en pre y postgrado, sea en modalidad diurna y vespertina, en formatos tanto presencial como online. Para ello, se trabaja con el capital humano ya existente y transversal a todas las escuelas y departamentos de la UVM. Se trabaja tanto en evaluación como en gestión en educación, teniendo en cuenta que la UVM es una universidad inclusiva, ayudando a problemas de la educación como la retención, la progresión y la eficacia académica de las y los estudiantes.

Luego de cumplir las metas y acciones presupuestadas para el 2021 y 2022, se pretende “escalar” el programa a un Centro de Investigación Educativa (CIE) que contenga las líneas de investigación del programa de IF y además las líneas de políticas públicas de educación escolar desde una perspectiva situada y la línea de didáctica de las ciencias naturales y matemáticas.

REFERENCIAS BIBLIOGRÁFICAS

- Neidhardt, F. C. (1997). A Michigan Tradition: Research and Undergraduate Education. Report to the Regents of the University of Michigan. www.research.umich.edu.
- Restrepo, G.B. (2003). Conceptos y aplicaciones de la Investigación Formativa y criterios para evaluar la investigación científica en sentido estricto. CNA. Bogotá, Colombia.
- Universidad Viña del Mar. (2019). Política de investigación de la Universidad Viña del Mar (<https://www.uvm.cl/wp-content/archivos/resolucion-08-2019-politica-investigacion.pdf>)

2

Creación e investigación formativa en la Universidad Viña del Mar Aproximaciones a la instalación, aportes y mecanismos de una noción emergente

Dr. José Rivera-Soto¹⁻², Alfredo González Sánchez¹

¹Dirección de Investigación y Postgrados

²jriviera@uvm.cl

RESUMEN

En el presente artículo, analizaremos el proceso de instalación y los avances en el posicionamiento institucional de la noción de Creación e Investigación Formativa (CIF) en la Universidad Viña del Mar. Para ello, revisaremos cómo se ha ido incorporando, de manera clara y definida, en las declaraciones institucionales de mayor alcance, con orientaciones respecto a sus características, función y vínculo con los objetivos y propósitos formativos de la Universidad, así como de la importancia que se le debe asignar y los mecanismos de que se dispone para que la CIF retroalimente el proceso de enseñanza y aprendizaje.

Palabras clave: Creación e investigación formativa; experiencias educativas situadas; instalación; mecanismos; Universidad Viña del Mar.

1. Introducción: Creación e investigación formativa, una noción emergente

¿Qué es la Creación e Investigación Formativa (CIF)?

La pregunta es acuciante en el contexto universitario, como muestra la constante presencia de dicho concepto en normativas y orientaciones de la Comisión Nacional de Acreditación (CNA), ente que establece criterios y estándares de la calidad de los procesos en la educación terciaria.

En ese marco, en mayo de 2021, el Programa de Investigación Formativa (PIF) de la Universidad Viña del Mar (UVM), a través de un breve cuestionario, consultó, entre otros tópicos, el grado de conocimiento sobre la noción entre los académicos¹ de la institución. La respuesta fue auspiciosa: la definición de investigación formativa -en este caso, la entregada por el especialista colombiano Bernardo Restrepo (2003), rescatada para el estudio en comento- era conocida por un 60% de los encuestados.

Esta definición -que fue un antecedente teórico clave en la elaboración del PIF-UVM-, indica que la investigación formativa es aquella que se hace entre estudiantes y docentes en el proceso de desarrollo del currículo de un programa y que es propio de la dinámica de la relación con el conocimiento que debe existir en todos los procesos académicos tanto en el aprendizaje, por parte de los alumnos, como en la renovación de la práctica pedagógica por parte de los docentes (p. 6).

Lo paradójico, es que existe consenso entre los investigadores del PIF y el directivo de la CNA de Colombia que elaboró la definición, que la misma resulta imprecisa, de bordes difusos respecto a lo que puede comprender este tipo específico de investigación.

Es más, el tema se hace aún más ambiguo cuando agregamos la palabra **creación** a este concepto.

Desde esta perspectiva, una nueva definición posible es la que leemos en el Documento de Trabajo Criterios de Evaluación para Carreras y Programas de Pregrado (s/f), de la CNA chilena, que conceptualiza la “Creación e Investigación formativa por el Cuerpo Docente” como la publicación de “trabajos académicos que impacten positivamente la teoría, la práctica y/o la enseñanza en forma consistente con la misión y visión institucional” (p. 26).

Ahora bien, más que arribar a una definición exacta, inequívoca de la CIF, en este texto intentaremos hacernos cargo de los factores que justifican su importancia.

A nuestro juicio, el principal de ellos, radica en que la CIF permite dar cuenta de experiencias educativas situadas, en contextos particulares, con perfiles de estudiantes concretos, algo que, en una sociedad que reclama el acceso a la formación terciaria como derecho, es de vital importancia, al posibilitar que las instituciones se hagan cargo de la heterogeneidad de quienes ingresan a las aulas universitarias en la actualidad.

¹ El presente texto, dada su extensión y en pos de mayor claridad, usará el genérico masculino.

Es decir, la CIF es la manera en que una Universidad, **teniendo en cuenta sus objetivos y propósitos formativos, instala mecanismos que aseguren la mejora continua en el proceso de enseñanza-aprendizaje, al integrar la situación educativa específica y concreta en que esta se desarrolla, ya sea mediante trabajos académicos y científicos** -publicados en revistas de corriente principal, libros o capítulos de libros, patentes, entre otros-, **o a través de materiales y recursos educativos, aplicaciones tecnológicas o vinculaciones académicas con grupos de estudio disciplinares.**

En este texto, analizaremos el caso particular de la Universidad Viña del Mar, revisando cómo la Creación e Investigación Formativa se ha favorecido en las declaraciones institucionales de mayor alcance y, luego, los mecanismos y aportes concretos que esta realiza a la formación de pre y postgrado.

2. La instalación de la CIF: su presencia en las declaraciones institucionales UVM

Como dijimos, la Comisión Nacional de Acreditación (CNA), aborda en distintas normativas la importancia de las actividades de creación e investigación formativa.

Podemos destacar, entre ellas, los Criterios de evaluación para la acreditación de carreras profesionales, carreras profesionales con licenciatura y programas de licenciatura (Res. Extensa N° DJ 009-4 - 2015), donde se refiere, en el criterio 10, a la “Creación e Investigación formativa por el Cuerpo Docente” (p. 13). De manera análoga, el Reglamento sobre Áreas de acreditación Institucional (Res. Extensa DJ N° 01 - 2013), indica como uno de los cinco aspectos clave a evaluar, la “Utilización de los procesos o resultados de la investigación para mejorar la calidad de la docencia impartida” (p. 5).

En el Documento de Trabajo Criterios de Evaluación para Carreras y Programas de Pregrado, citado previamente, la CNA entrega mayor detalle sobre lo que entenderá por CIF. Tras expresar que las unidades académicas deben promover “que entre sus docentes se publiquen trabajos académicos que impacten positivamente la teoría, la práctica y/o la enseñanza en forma consistente con la misión y visión institucional” (p. 26), el documento divide en cuatro los aspectos a considerar, enumerándolos como sigue:

10. a La unidad académica que imparte la carrera o programa promueve que sus docentes generen, publiquen o expongan trabajos académicos –nuevos conocimientos o aplicaciones de conocimientos o innovaciones en las metodologías y/o materiales de enseñanza– conducentes a mejorar la docencia en la consecución del perfil de egreso.

10. b Los docentes que en su conjunto constituyen el núcleo de alta permanencia de la carrera o programa, han publicado en los últimos 5 años materiales educativos que contribuyen a la enseñanza de la carrera o programa y/o resultados de investigaciones que han hecho avanzar la teoría y/o la práctica.

Las investigaciones y/o materiales educativos pueden corresponder a:

a) Investigación básica que genera y comunica nuevos conocimientos en publicaciones académicas ya sea en la forma de artículos, capítulos de libros, libros o materiales de enseñanza en las disciplinas propias de la carrera o programa.

b) Aplicaciones que desarrollan nuevas tecnologías, procesos, herramientas y usos haciendo avanzar nuevos métodos de trabajo que tienen por objeto impactar la práctica en la disciplina.

c) Materiales de enseñanza y aprendizaje que desarrollan y hacen avanzar nuevas interpretaciones, y contenidos de enseñanza y metodologías que impactan el aprendizaje. Los aportes en este acápite normalmente tienen por objeto fortalecer la enseñanza de la carrera o programa.

10.c La carrera o programa monitorea y evalúa los trabajos académicos en función del cumplimiento de los objetivos declarados.

10.d La carrera o programa mantiene relaciones académicas con centros, grupos, redes o programas dedicados a la investigación en el campo disciplinar (p. 27-28).

Teniendo a la vista las recomendaciones de la CNA, junto a la experiencia de los académicos de pre y postgrado, y la impronta y sello de la propia Universidad, durante los últimos años la creación e investigación formativa se ha posicionado en la UVM como una actividad que produce conocimiento a la par que fortalece la calidad de la docencia. Esto se verifica en la presencia del concepto en cada una de las declaraciones que guían el quehacer institucional en el mediano y largo plazo. Veamos.

Como primer elemento, podemos consultar la Política de Investigación vigente (Resolución N°08/2019, Universidad Viña del Mar), que cuenta entre sus objetivos visibilizar la investigación realizada por la Universidad, así como favorecer y valorar la investigación formativa. Luego, al determinar cómo será entendida la productividad en investigación formativa, la Política abre el concepto a la creación e investigación formativa, indicando que se hará con publicaciones en revistas indexadas, como también a través de la publicación de materiales de enseñanza y la generación de aplicaciones tecnológicas, desarrolladas a partir de prácticas docentes, así como la participación en redes de reflexión y trabajo colaborativo con otros investigadores en esta línea (p. 3).

Algo similar ocurre con la Política de Aseguramiento de la Calidad (Resolución N°05/2021, Universidad Viña del Mar) que, al describir la función misional de investigación, creación e innovación, tiene entre sus indicadores de resultados para evaluar el impacto y la calidad de la investigación y la generación de conocimiento en la Universidad el incremento en el número de publicaciones, productos y materiales de creación e investigación formativa.

Como señalaban los documentos CNA, es indispensable que la CIF se dé al interior del marco mandatorio del Proyecto Institucional y el Modelo Educativo, algo que también se ha intencionado.

En cuanto al Proyecto Institucional, este reconoce de manera explícita el propósito de incorporar y potenciar la CIF al afirmar:

En consistencia con nuestra identidad, tanto en el nivel de pregrado como en el posgrado, se favorece la creación e investigación formativa, fomentando así, la generación de conocimiento propia de los procesos académicos, que impacte a los estudiantes en su aprendizaje, así como también al académico en una mejor práctica pedagógica, conduciendo, al mismo tiempo, a resultados que aseguren un adecuado despliegue del Proyecto Institucional (p. 5).

En el Modelo Educativo, en tanto, entendido como la concreción del Proyecto Institucional en el ámbito académico y formativo, se declara:

Nuestra convicción es que la investigación científica y disciplinar, tal como la de carácter formativo, deben colaborar positivamente en los procesos de enseñanza aprendizaje, así como ser pertinentes y significativas para el contexto donde se desarrollan (p. 12). (...) Es mediante la investigación que le ofrecemos un rol activo al estudiante, respecto a los conocimientos que componen la disciplina en la que se forma, favoreciendo las competencias necesarias para la indagación científica. (...)

La investigación, de igual forma, debe integrarse en el proceso mismo de enseñanza aprendizaje, desde las distintas modalidades que ofrece la investigación formativa, relevando la importancia de una docencia consciente de sus posibilidades de mejora (p. 13).

Esto enseña que, en suma, la creación e investigación formativa en la UVM coadyuva a definir el quehacer misional, plasmado en su Proyecto y Modelo, imbricando la CIF a los procesos de enseñanza y aprendizaje, a partir de las distintas modalidades en que esta se despliega.

Como dijimos al inicio, el concepto de creación e investigación formativa ha tenido un interesante correlato al nivel de las declaraciones institucionales que orientan el quehacer de la UVM. Con ello, la Universidad se hace cargo de las situaciones educativas de un modo situado, en su contexto particular, donde “su pertinencia viene dada por los objetivos curriculares o los propósitos de formación del programa académico dentro del cual se desarrolla la investigación formativa” (Parra Moreno, 2004, p. 74), y donde, además, “su significado académico y la novedad de sus resultados tienen] una importancia local” (Hernández, 2003, p. 185), consistentes con la misión y visión institucional.

3. Investigación que retroalimenta los procesos formativos

Para la Universidad Viña del Mar, la investigación, desarrollo e innovación constituyen actividades relevantes para la mejora continua de los procesos formativos, considerando sus resultados como un aporte permanente a la docencia de pregrado. Con ello se persigue que las acciones de investigación retroalimenten los procesos de enseñanza-aprendizaje, de manera que sean medibles y permitan evaluar los grados y calidad de esa incidencia.

Con ese objetivo, se han definido una serie de mecanismos en los que la investigación colabora a la mejora de la docencia, determinando, para cada uno de ellos, las métricas para evaluar los aportes. Dichas métricas se basan en los indicadores de resultado que ha definido la propia Universidad, y que se presentan en la “Matriz de calidad de la función misional formación de pre y postgrado”, de la Política de Aseguramiento de la Calidad UVM.

Sin duda, los principales mecanismos son los ingentes fondos que la UVM destina, anualmente, para el desarrollo de la creación e investigación formativa, a la manera de un incentivo formal institucional para la realización de la CIF, algo que viene haciendo sistemáticamente desde hace una década a través de dos concursos internos.

Ambos concursos, organizados por la Vicerrectoría Académica, son el Fondo de Creación e Investigación Formativa (FCIF), dependiente de la Dirección General de Docencia, específicamente, alojado en la Subdirección de Innovación Educativa, y los Fondos Internos de Investigación (FII), en su Línea Priorizada de Investigación Formativa, dependiente de la Dirección de Investigación y Postgrados.

El primero de ellos, tuvo su primera versión en 2011, bajo el nombre de Fondo de Mejoramiento e Innovación Docente (FID). En 2020, el concurso cambia de nombre, en sintonía con la importancia que la Universidad fue asignándole a la CIF, como hemos comentado.

En estos 10 años, el FCIF ha financiado más de cien proyectos de innovación e investigación educativa de los académicos UVM, con un marcado énfasis en la reflexión y sistematización de las experiencias docentes, así como en la innovación y producción de conocimiento relevante para los procesos de enseñanza-aprendizaje.

Este Fondo promueve cuatro líneas de acción en que los postulantes deben inscribir sus propuestas, a saber:

- Efectividad de la docencia virtual;
- Alineamiento constructivo;
- Evaluación auténtica del aprendizaje; y
- Docencia inclusiva.

El segundo mecanismo es el Fondo Interno de investigación (FII), que se aboca, específicamente, al trabajo científico y la producción de conocimiento en distintas áreas del saber, alineado también con la misión y visión de la Universidad.

El FII fue creado en 2014, junto con la aparición de la primera Política de Investigación (Resolución N°05/2014, Universidad Viña del Mar), y anualmente financia más de 15 proyectos en alguna de sus tres líneas:

- Línea Priorizadas de Investigación (LPI), que posee cuatro temáticas de carácter misional: Investigación formativa, Inclusión social, Desarrollo Territorial y Salud y bienestar;
- Línea de Ciencia, Tecnología y Conocimiento (CTC), con temáticas determinadas por la curiosidad del investigador; y
- Línea de Estudiantes, Titulados y Graduados, de iniciación en investigación para cursantes y egresados de nuestra casa de estudios.

El acento en la creación e investigación formativa que tiene el FII, puede observarse tanto en la Línea priorizada que lo vincula como un área de relevancia para la institución, como en la cantidad de proyectos que todos los años son postulados a temáticas vinculadas a la CIF en la línea de CTC.

Empero, estos no son los únicos mecanismos de que dispone la Universidad para fomentar la CIF en el cuerpo académico. A modo de resumen, describimos la totalidad de los mismos, a partir de cuatro dimensiones genéricas, detalladas en la siguiente tabla:

Dimensiones	Mecanismos
1. Investigación educativa	1.1 Fondo de Creación e Investigación Formativa, FCIF
	1.2 Fondo Interno de Investigación, LPI, Investigación formativa
	1.3 Programa de Investigación Formativa, PIF-UVM
2. Recursos Pedagógicos	2.1 Materiales de enseñanza, aplicaciones de nuevas tecnologías, procesos y herramientas
	2.2 Repositorio institucional Dspace
	2.3 Ediciones UVM
3. Investigación científica y producción de conocimiento	3.1 Fondos Internos de Investigación
	3.2 Productividad científica del cuerpo académico en Programas y Syllabus de asignatura
4. Relaciones Académicas	4.1 Relaciones Formales con Centros, Programas y Redes
	4.2 Congresos, Coloquios, Seminarios y Otros

Durante 2021, la Universidad ha emprendido la tarea de medir, transversalmente, los resultados e impacto que cada uno de estos mecanismos tienen en la docencia de pre y postgrado. Para ello, ha elaborado una serie de instrumentos de recolección de datos, que ha venido aplicando para conocer el grado de retroalimentación y aporte que existe entre ambas funciones misionales, la formación y la investigación.

Esperamos esa indagación ofrezca hallazgos positivos que incentiven a los académicos y la propia institución a perseverar en el trabajo y desarrollo de la creación e investigación formativa en el futuro.

REFERENCIAS BIBLIOGRÁFICAS

- Restrepo, G.B. (2003). Conceptos y aplicaciones de la Investigación Formativa y criterios para evaluar la investigación científica en sentido estricto. CNA. Bogotá, Colombia. Disponible en: <https://www.epn.edu.ec/wp-content/uploads/2017/03/Investigaci%C3%B3n-Formativa-Colombia.pdf>
- Comisión Nacional de Acreditación, CNA-Chile. (s/f). Criterios de Evaluación para Carreras y Programas de Pregrado. Documento de Trabajo. Disponible en: <https://www.cnachile.cl/noticias/SiteAssets/Paginas/consultapublica/CRITERIOS%20DE%20EVALUACI%C3%93N%20PARA%20CARRERAS%20Y%20PROGRAMAS%20DE%20PREGRADO.pdf>
- Parra Moreno, C. (2004). Apuntes sobre la investigación formativa. Educación y Educadores, vol. 7. Universidad de La Sabana. Chía, Colombia, p. 74. Disponible en: <https://educacionyeducadores.unisabana.edu.co/index.php/eye/article/view/549/642>
- Hernández, C. A. (2003). Investigación e Investigación Formativa. Nómadas, núm. 18. Universidad Central Colombia. Bogotá, p. 185. Disponible en: <https://www.redalyc.org/pdf/1051/105117890018.pdf>
- Universidad Viña del Mar. (2021). Política de Aseguramiento de la Calidad de la Universidad Viña del Mar. Disponible en: <https://www.uvm.cl/wp-content/archivos/politica-aseguramiento-calidad.pdf?v=2021>
- Universidad Viña del Mar. (2019). Política de investigación de la Universidad Viña del Mar. Disponible en: <https://www.uvm.cl/wp-content/archivos/resolucion-082019-politica-investigacion.pdf>
- Universidad Viña del Mar. (2021). Proyecto Institucional de la Universidad Viña del Mar. Disponible en: <https://www.uvm.cl/wp-content/archivos/proyecto-institucional.pdf>
- Universidad Viña del Mar (2021). Modelo Educativo de la Universidad Viña del Mar. Disponible en: <https://www.uvm.cl/wp-content/archivos/modelo-educativo.pdf>

3

Análisis de las historias de clase de profesores que reflexionan para una educación transformadora en ciencias: estructura y cambios en el tiempo

Camilo Henríquez Miranda¹, Corina González-Weil²,
Paulina Bravo González²

¹Universidad Viña del Mar. Escuela de Ciencias

²Pontificia Universidad Católica de Valparaíso

¹chenriquez@uvm.cl

RESUMEN

La educación en ciencias requiere cambios importantes acordes a los tiempos actuales, siendo la reflexión docente un proceso crucial en el desarrollo profesional de docentes de esta disciplina. Si bien las instancias formales que promueven la reflexión docente son muy escasas en el sistema educativo chileno, existen comunidades de aprendizaje informales, en donde se fomenta el reflexionar y compartir experiencias pedagógicas. Una de las maneras de compartir las experiencias profesionales y el conocimiento que se genera a partir de la reflexión es la escritura de relatos sobre lo que ocurre en el aula. Este estudio analiza la estructura y transformación de 48 “historias de clase” (HdC) que la comunidad de aprendizaje Profesores Reflexionando por una Educación Transformadora en Ciencias (PRETeC, por su sigla en español) ha compartido y discutido entre los años 2013 y 2017. Los resultados muestran la presencia de una estructura base y evidencia una transformación a través del tiempo, dando luces de cómo las prácticas relatadas por los docentes participantes del grupo van cambiando desde una enseñanza centrada en el docente a una centrada en los estudiantes. Por otra parte, se analizó la presencia y desarrollo de elementos del Conocimiento Pedagógico del Contenido (PCK) en la historia de clase, encontrando una relación entre el cambio en estos elementos y los cambios en la estructura de las historias, en donde a medida que pasa el tiempo, los docentes reflexionan más en torno a sus estudiantes y las consecuencias de sus prácticas en el aula. Des esta forma podemos indicar

que el análisis de las historias de clases de una comunidad de aprendizaje que comparte sus historias de clase de forma sistemática ha permitido visualizar un desarrollo tanto en aspectos del PCK como en cambios en la práctica de quienes escriben las historias, cambios que a su vez repercuten en su desarrollo profesional docente.

Palabras clave: Reflexión, PCK, narrativas, trabajo colaborativo, desarrollo profesional.

Introducción

Para los docentes de ciencias hoy en día, es difícil encontrar instancias para reflexionar, no obstante, el reflexionar sobre la práctica pedagógica es considerado uno de los aspectos más relevantes dentro de las habilidades, conocimientos y aptitudes que debería tener un profesor de ciencias. Una de las maneras de promover la reflexión entre docentes, es mediante su participación en comunidades de aprendizaje entendidas como grupos de profesores que comparten y cuestionan críticamente su práctica de una manera continua, reflexiva, colaborativa y orientada al aprendizaje, para promover su crecimiento y sus habilidades (Vangrieken et al, 2015).

Por su parte, el uso de narrativas (Le Fevre, 2011) es una de las maneras de compartir y sistematizar las experiencias de aula, constituyendo una valiosa herramienta de reflexión que incide directamente en el desarrollo profesional de los profesores participantes, ya que les hace ser conscientes de las diferentes acciones que realizan dentro de la sala de clases, situación que abre el camino hacia la transformación de las prácticas (González-Weil et al, 2014). A su vez, el uso de narrativas por un tiempo prolongado puede en sí mismo servir de instrumento de análisis para investigar de qué manera se desarrolla profesionalmente el docente, cómo cambia su conocimiento acerca de enseñar ciencias (PCK) y cómo cambia su práctica.

El presente estudio analiza las narrativas presentadas y compartidas en la comunidad Profesores Reflexionando por una Educación Transformadora en Ciencias (PRETeC, por su sigla en español) entre los años 2013 y 2017. Estas narrativas son denominadas por el grupo como “historias de clase” (HDC), y corresponden a una narración escrita de una sesión de clases. Dado la larga trayectoria del grupo (8 años, 5 compartiendo historias de clase) y la gran cantidad de historias de clase compartidas (48), surgen los siguientes objetivos de investigación: (1) Describir la estructura de la historia de clase y los cambios de ésta a través de los años. (2) Identificar los elementos de PCK presentes en el aprendizaje que declaran haber desarrollado los docentes mediante la escritura de la historia de clase y su cambio en el tiempo. (3) Establecer las características de la transformación de la práctica de los autores de las HDC, mediante la relación de la evolución de los elementos de la HDC y los aspectos identificados del PCK.

Desarrollo

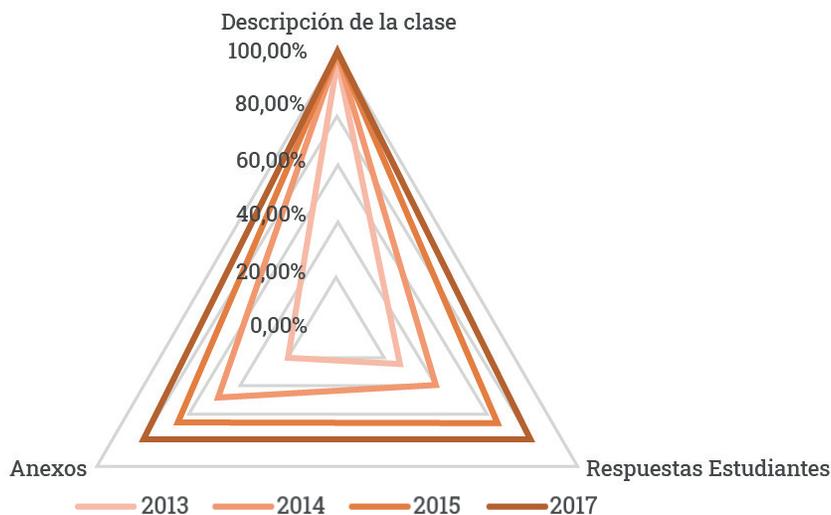


Figura 1: Frecuencia relativa porcentual de elementos que forman el desarrollo de la clase en una historia de clase

La estructura de la historia de clase: Se identificaron tres grandes apartados: Contextualización, Desarrollo de la clase y Conclusión. La contextualización corresponde al momento inicial de la HDC en el cual, quien narra la historia entrega diferentes detalles, entregando un marco de referencia general. El desarrollo de la clase corresponde al momento en el que se describe la clase propiamente tal y se entrega información que permita comprender lo que sucedió en ella. En la conclusión, se plasman algunas ideas a modo de cierre de la historia. Cada apartado se compone de diferentes elementos:

- (1) Contextualización: Título; Contenido a trabajar: Contextualización del curso; Contextualización de la sesión; Objetivo de la clase. La presencia de estos elementos cambia a través de los años en los que se escriben historias de clase cambiando también sus características.
- (2) Desarrollo de la Clase: Descripción de la clase; Respuestas de los estudiantes; Anexos. Según la figura 1, se observa un aumento en la presencia de “respuestas de los estudiantes (25% en 2013 a 78% en 2017)” y de “anexos” (20% en 2013 a 78% en 2017), en los cuales los profesores van incorporando evidencias de su clase.
- (3) Conclusiones: Aprendizaje del autor (55% en 2013; 78% en 2017); Aprendizaje del estudiante (50% en 2013; 67% en 2017). Del mismo modo que en los dos momentos anteriores se puede observar una variación de la presencia de estos elementos por año de escritura, observándose un aumento en ambas categorías.

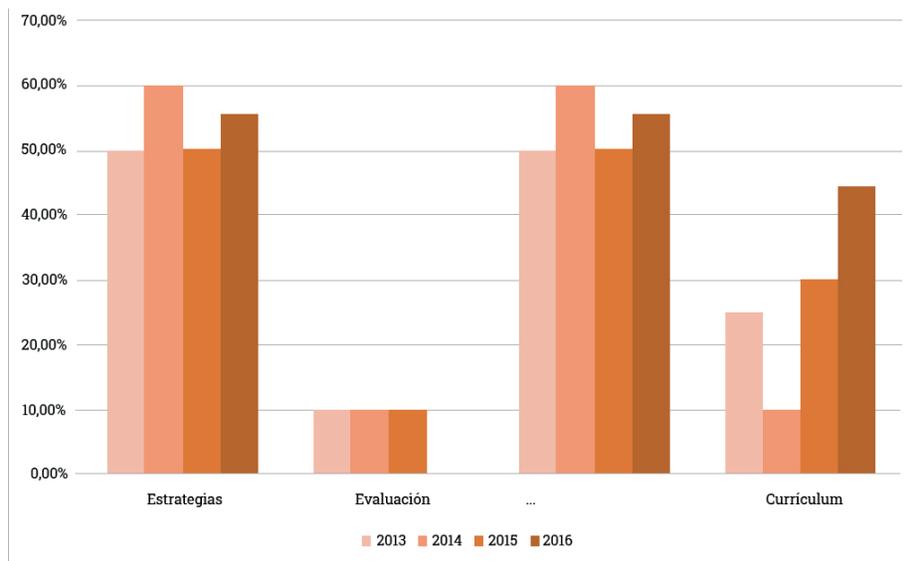


Figura 2: Frecuencia relativa porcentual de los elementos de PCK presentes en el apartado aprendizaje del autor

(4) Elementos del conocimiento pedagógico del contenido (PCK) presentes en el apartado aprendizaje del profesor: Siguiendo a Magnusson et al (1999), se consideraron 4 elementos: conocimiento sobre estrategias, sobre evaluación, sobre la comprensión de los estudiantes, sobre el curriculum. Mientras las dos primeras se mantienen, las dos últimas categorías aumentan fuertemente a lo largo de los años (figura 2).

Sobre la relación entre los cambios en la estructura de las HDC y los elementos de PCK declarados como aprendizajes por los docentes, se observa una gran coincidencia, por cuanto a medida que va aumentando el conocimiento sobre los estudiantes (elemento de PCK), va aumentando la mención que se hace de las respuestas de los estudiantes (respuesta estudiantes), su aprendizaje, (aprendizaje estudiantes) y las evidencias de ello (anexos). Lo anterior, se relaciona también con una mayor presencia del apartado “aprendizaje del autor”, lo que indica que los docentes del grupo se van haciendo más conscientes de su propio aprendizaje.

Conclusiones y/o recomendaciones

La estructura de las HDC cambió drásticamente entre los años 2013 a 2017, generándose un corrimiento desde un enfoque más centrado en la enseñanza a uno centrada en el aprendizaje. Lo anterior coincide con lo señalado por Vangrieken et al (2015) como una de las consecuencias del trabajo colaborativo docente, que en este caso está dado por la comunidad de aprendizaje PRETeC. A su vez, el aumento progresivo del elemento “Aprendizaje del autor” representa una mayor reflexión acerca de la clase, lo cual coincide con la idea de que compartir experiencias pedagógicas a través de la narrativa influye de manera significativa como estímulo a reflexión.

REFERENCIAS BIBLIOGRÁFICAS

- González-Weil, C., Waring, M. G., Albayay, G. A., González, P. B., Tapia, E. S., Cisternas, D. A., ... & Valenzuela, J. S. (2014). Principios de Desarrollo Profesional Docente contruidos por y para Profesores de Ciencia: una propuesta sustentable que emerge desde la indagación de las propias prácticas. *Estudios Pedagógicos*, 40, 105-126.
- Le Fevre, D. M. (2011). Creating and facilitating a teacher education curriculum using preservice teachers' autobiographical stories. *Teaching and Teacher Education*, 4, 779-787.
- Magnusson, S., Krajcik, J., & Borko, H. (1999). Nature, Sources and development of Pedagogical Content Knowledge for Science Teaching. In J. Gess-Newsome & N. Lederman (Eds.), *PCK and Science Education* (pp. 95-132). Netherlands.
- Vangrieken, K., Dochy, F., Raes, E., & Kyndt, E. (2015). Teacher collaboration: A systematic review. *Educational Research Review*, 15, 17-40.

SEGUNDA PARTE

LA INVESTIGACIÓN DESDE LA ÓPTICA MULTI E INTERDISCIPLINAR

1

Paradigmas de la didáctica de las Matemáticas y su aplicación en particular al objeto matemático de optimización

Silvia Soledad López

*Escuela de Ciencias
silvia.lopez@uvm.cl*

RESUMEN

La didáctica de la matemática estudia el proceso de transmisión de conocimientos del profesor al alumno, pero también cómo el alumno se apropia y reconstruye estos conocimientos. Estos procesos de adquisición de conocimientos son diferentes en función de los conocimientos impartidos. Para cada concepto matemático, es importante conocer cómo se ha construido históricamente y los problemas que permite resolver, para comprender los obstáculos que planteará su enseñanza. De hecho, la didáctica de las matemáticas también busca crear situaciones de enseñanza adecuadas que permitan superar estos obstáculos y transmitir los conocimientos.

Brousseau plantea la hipótesis de que para cada noción que se va a enseñar, hay al menos una situación fundamental que se propondrá a los alumnos. Esta situación es problemática. La noción que hay que adquirir es la herramienta privilegiada para resolverlo. El alumno construirá la noción resolviendo el problema. Haciendo un análisis epistemológico y didáctico del concepto, podemos estudiar los obstáculos que encontrará el alumno y variar esta situación fundamental para ayudarlo a superar estos obstáculos. Para el alumno, lo que está en juego es resolver el problema y demostrar la validez de su solución.

La didáctica de las matemáticas se presenta así como la ciencia de las condiciones de difusión de los conocimientos matemáticos útiles para los hombres y sus instituciones. Con el objetivo de ejemplificar lo que se plantea desde la didáctica de las matemáticas para el estudio científico en este campo, este escrito presenta algunos modelos teóricos, y en particular una investigación que fue realizada con estudiantes de la UVM en los años

2018-2019, que revela en concreto un estudio sobre la enseñanza y aprendizaje de la matemática en el dominio del análisis, focalizada en el estudio de la optimización con funciones de una y dos variables.

Palabras clave: Marcos teóricos, paradigmas, didáctica de la matemática, optimización.

Introducción

En Francia, el estudio de la didáctica de la matemática se inicia en el año 1970 con la creación de los primeros IREM (Instituto para la investigación de la enseñanza de la matemática). Con sus principales autores Brousseau (autor de la Teoría de las Situaciones Didácticas y de numerosos conceptos didácticos teóricos (TSD)), Chevallard, (autor de la teoría de la transposición didáctica y la Teoría Antropológica de la Didáctica (TAD)) y Vergnaud (autor de la teoría de los campos conceptuales, cuyas nociones ejes son: campo conceptual, esquema y competencia).

Como característica de esta línea puede citarse el interés por establecer un marco teórico original, desarrollando sus propios conceptos y métodos y considerando las situaciones de enseñanza y aprendizaje globalmente. Los modelos desarrollados comprenden las dimensiones epistemológicas, sociales y cognitivas y tratan de tener en cuenta la complejidad de las interacciones entre el saber, los alumnos y el profesor, dentro del contexto particular de la clase.

El primer concepto creado por G. Brousseau, que formó parte de los demás desarrollos, es el de la Teoría de las Situaciones, formulada en su primera fase a principios de los setenta, desarrollada en una segunda fase hasta la publicación de la tesis de Brousseau y seguida por los aportes de Chevallard (1990) en términos de instituciones y de las relaciones con el saber.

Brousseau (1986) establece que:

“La didactique des mathématiques” étudie les activités didactiques, c’est-à-dire les activités qui ont pour objet l’enseignement, évidemment dans ce qu’elles ont de spécifique aux mathématiques.

Les résultats, dans ce domaine, sont de plus en plus nombreux, ils portent sur les comportements cognitifs des élèves, mais aussi sur les types de situations mis en œuvre pour les leur enseigner et sur tous les phénomènes auxquels la communication du savoir donne lieu. La production ou l’amélioration de moyens d’enseignement trouve dans ces résultats bien plus que des objectifs ou des moyens d’évaluation, elle y trouve un appui théorique, des explications, des moyens de prévision et d’analyse, des suggestions, voire des dispositifs et des méthodes. (Brousseau, 1986, p. 35)

En lo que respecta a este escrito mostramos cómo un modelo teórico de la didáctica de la matemática, es utilizado para un estudio científico en este campo de investigación. Nos referimos al objeto matemático de la optimización, la cual se enseña inicialmente centrándose en las funciones de una variable a los estudiantes de secundaria, con el criterio de la primera derivada para los puntos críticos y la monotonía. El criterio de la segunda derivada también puede utilizarse para determinar la naturaleza de los puntos críticos. Sin embargo, lo que se explica para funciones de una variable, con el uso de la monotonía, no puede extenderse a funciones de dos variables y la explicación del criterio es insuficiente para los casos en que el determinante de la matriz Hessiana es igual a cero. En la universidad predomina el trabajo algebraico con el uso del criterio; no se utilizan herramientas dinámicas, como GeoGebra. Por otro lado, es difícil dar sentido a la optimización, y los procesos cognitivos son pobres. En consecuencia, podemos plantear la siguiente hipótesis: los estudiantes universitarios no comprenden la optimización. Para abordar esto, se diseñó una situación didáctica y se implementó con estudiantes de ingeniería de la UVM. En particular, se consideró el proceso de visualización a través del registro gráfico, utilizando la herramienta dinámica GeoGebra, para mejorar la comprensión local de problemas de optimización con aproximaciones de Taylor de primer y segundo orden. El objetivo es desarrollar una significativa aproximación a la optimización que pueda extenderse de las funciones de una variable a las de dos variables.

Desarrollo

En primer lugar, nos referiremos a la teoría del espacio del trabajo matemático (ETM), los paradigmas del análisis y las perspectivas de localización, y a continuación estableceremos nuestra metodología.

La teoría del ETM

El objetivo de la teoría ETM es describir, analizar, diseñar y comprender el trabajo matemático propuesto a un individuo en una institución y realizado por éste.

Para caracterizar el trabajo matemático, la teoría considera aspectos epistemológicos y cognitivos que se organizan en dos planos: el plano epistemológico y el plano cognitivo. En cada uno de estos planos se consideran tres componentes, organizados en tres dimensiones, también denominados génesis, que se consideran esenciales en el trabajo matemático (figura 1):

- La Génesis semiótica que considera la visualización como un proceso cognitivo donde un individuo da significado a los signos matemáticos (considerados, por ejemplo, en registros de representación semiótica) de la componente representamen;

- La Génesis instrumental en la que un individuo transforma un artefacto (que puede ser material, digital o simbólico) en un instrumento para utilizarlo en un proceso cognitivo de construcción;

- La Génesis discursiva en la que un individuo utiliza el conocimiento matemático (definiciones, teoremas, propiedades,...) del referencial teórico en el proceso cognitivo de comprobación (en sentido amplio).

Obviamente, estas tres dimensiones están interrelacionadas. Se necesitan conocimientos para poder visualizar los signos matemáticos o utilizar eficazmente un artefacto; la demostración utiliza representaciones semióticas de los objetos matemáticos; la construcción crea nuevos signos matemáticos, etc. La teoría del ETM pretende comprender el papel de cada una de estas dimensiones y componentes en un sistema integrado que da lugar al trabajo matemático.

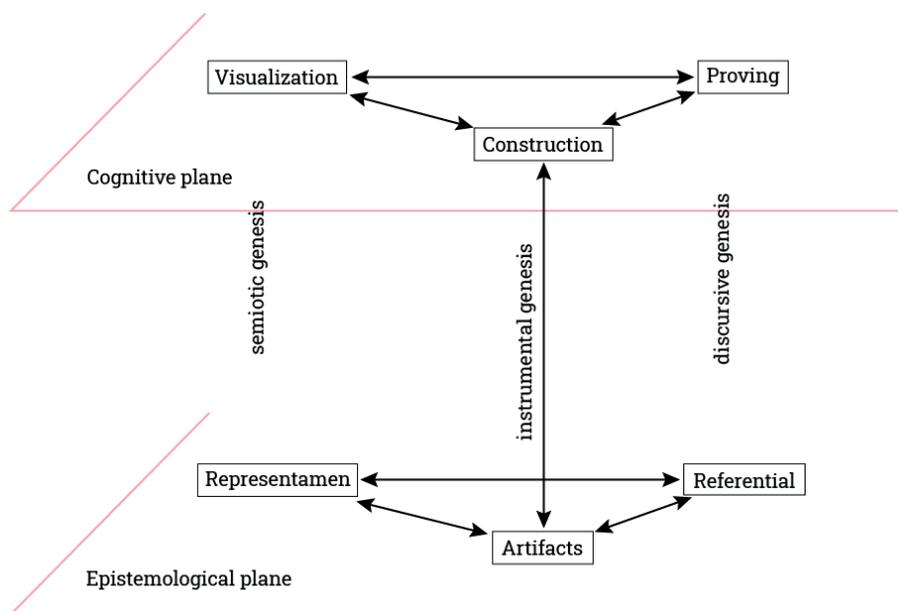


Figura 1: El modelo de los ETM (Kuzniak et al., 2016, p725)

La articulación y las interrelaciones entre los diferentes géneros se tienen en cuenta a través de planos verticales. Cada uno de ellos destaca el papel preponderante de dos géneros: el plano semiótico-instrumental [*Sem-Ins*], el plano instrumental-discursivo [*Ins-Dis*] y el plano semiótico-discursivo [*Sem-Dis*].

Los artefactos desempeñan un papel crucial en nuestro estudio. Estos pueden ser de tres tipos:

- Materiales, como brújulas o transportadores, que no se tratan aquí;
- Digitales, como las calculadoras o los programas informáticos dinámicos como GeoGebra, estos últimos juegan un gran papel en nuestro estudio;
- Artefactos simbólicos, es decir, rutinas o algoritmos que se utilizan en el proceso de construcción sin hacer ninguna conexión con la dimensión discursiva donde se justifican.

Los artefactos simbólicos son importantes en nuestro estudio, como las reglas de cálculo de la derivada y el criterio de la segunda derivada (si $f'(a)=0$ y $f''(a)>0$, entonces a es un punto mínimo de f). Los artefactos simbólicos son muy útiles en el trabajo matemático, debido a su eficacia. Sin embargo, es importante que el individuo sea capaz de reflexionar y de dar sentido al uso de los artefactos simbólicos en algunos casos, como cuando el artefacto no da respuesta o da una respuesta extraña. En nuestro estudio sobre la enseñanza y el aprendizaje de la optimización, nos centraremos en el criterio de la(s) segunda(s) derivada(s) (parcial(es)), especialmente en el caso donde el determinante de la matriz Hessiana es igual a cero.

Sin embargo, también es posible preguntarse si la enseñanza institucional promueve la enseñanza (del criterio de la segunda derivada) de la aparición del sentido. Si sólo se pide a los alumnos que utilicen un artefacto simbólico en una dimensión instrumental, sin reflexionar sobre él y sin hacer ninguna conexión con las otras dos dimensiones, consideramos que esta institución no organiza la enseñanza para dar sentido a este artefacto simbólico. Por lo tanto, se prestará especial atención a lo que se denomina un trabajo completo, es decir, un trabajo que permita activar las tres génesis y planos de forma coordinada.

Paradigmas del análisis

La noción de paradigma es un elemento clave de la teoría del ETM que tiene por objeto comprender el trabajo realizado por un individuo, su orientación, las elecciones realizadas por una institución y, a continuación, caracterizar el trabajo aceptado por una comunidad o una institución. Dependen de un campo matemático, lo que permite tener en cuenta la especificidad del trabajo matemático en cada campo. Montoya y Vivier (2016) definieron los tres paradigmas del análisis:

- *ANI* que permite interpretaciones, de manera implícita, sobre la base de la geometría, el cálculo aritmético o el mundo real;

- AN2 donde las reglas del cálculo se definen y aplican independientemente de la reflexión, y de la existencia y naturaleza de los objetos matemáticos involucrados;

- AN3 se caracteriza por un trabajo, posiblemente topológico, sobre la aproximación (desigualdades, despreciabilidad,...) y las vecindades; la existencia y la naturaleza de los objetos matemáticos se establecen teóricamente.

Por ejemplo, cuando se utiliza el artefacto simbólico del criterio de la segunda derivada, se trabaja en el paradigma AN2. Al visualizar un gráfico, o una tabla de valores, llegar a la conclusión de que la función tiene un punto mínimo es un trabajo en el paradigma AN1. En un caso donde el determinante de la matriz Hessiana es igual a cero, utilizar el hecho de que $f''(a)$ es distinto de cero en una aproximación local para concluir que a no es un punto óptimo es un trabajo en el paradigma AN3 (aunque puede convertirse en un artefacto simbólico).

Por lo tanto, encontrar el significado también depende del cambio de paradigma, lo que implica una articulación entre los diferentes paradigmas. Sin embargo, la mayoría de las instituciones de enseñanza hacen hincapié en el paradigma AN2, ya que da valores precisos y es bastante fácil de evaluar; el AN1 se utiliza, pero, a menudo, sólo para ilustrar porque se considera que no es lo suficientemente matemático o no es preciso; el AN3 suele ser difícil para los estudiantes y se enseña en análisis avanzados.

Perspectivas de localización

Ampliando el trabajo de Vandebrouck sobre las funciones (Vandebrouck, 2011), Montoya, Páez, Vandebrouck y Vivier (2018) afirman que la adopción de diferentes perspectivas en la visualización (no icónica) de objetos matemáticos es una actividad cognitiva crucial en el campo del análisis matemático. Destacan tres perspectivas de localización:

- La perspectiva puntual (PP) implica objetos y nociones vistos como números reales aislados. Ejemplos: el valor $f(x_0)$ de una función f , el punto $M_0(x_0, f(x_0))$ en su curva, [...]

- La perspectiva global (PG) se refiere a los objetos vistos como un todo, en su totalidad, o al menos una gran parte de ellos. [...]. Por ejemplo, podemos considerar una función cuadrática como una función cuya curva es una parábola. [...]

- La perspectiva local (LP) es crucial en el campo del Análisis Real, [...]. Se trata de propiedades que se definen a partir de vecindades [...]. La precisión que proporciona el registro o el artefacto en cuestión es importante porque puede determinar la precisión de la vecindad. Por ejemplo, con GeoGebra, el máximo acercamiento a una curva se consigue cuando, visualmente, percibimos una línea recta. (Montoya et al. 2018, p151).

La noción de Deconstrucción con Perspectivas de Localización (DWLP): “DWLP se refiere a la actividad cognitiva que consiste en adoptar al menos una de las tres perspectivas de localización (puntual, local o global)” (Montoya et al. 2018, p151).

La visualización de los signos matemáticos nunca es un proceso puramente semiótico, ya que se relaciona con los conocimientos y nociones de que dispone un sujeto. Por lo tanto, es un proceso importante en la adquisición de significado. Para desarrollar el significado en el campo del análisis matemático, es importante adquirir flexibilidad entre diferentes perspectivas.

Metodología

Esta investigación ha consistido, en primer lugar, en definir cómo se enseña la optimización. Nuestra hipótesis es que esta enseñanza se reduce al uso de artefactos simbólicos (dimensión instrumental), con el criterio de la segunda derivada para las funciones de \mathbf{R} a \mathbf{R} y el criterio de las segundas derivadas parciales para las funciones de \mathbf{R}^2 a \mathbf{R} . Este trabajo matemático deja de lado los otros elementos que faltan en el plano epistemológico y cognitivo y se propone esencialmente en el paradigma AN2.

Para validar esta hipótesis, se analizaron los planes de estudio y libros de texto que se relacionan con la optimización de funciones utilizados por los estudiantes de ingeniería chilenos al inicio de la universidad. Luego, se analizó una prueba propuesta durante el segundo año de universidad para la optimización de funciones de dos variables reales.

Después de validar en tres universidades chilenas (dos públicas y una privada) que no promueven la aparición del sentido en este estudio y dado que éste no aparece de manera espontánea, se diseña una secuencia de enseñanza considerando una DWLP con la visualización local en la gráfica de funciones utilizando un software (GeoGebra) y basándose matemáticamente en las aproximaciones de Taylor de primer y segundo orden. El propósito es realizar un trabajo matemático completo que articule los paradigmas AN1 y AN2 (y elementos de AN3). El objetivo es identificar en la enseñanza de la optimización de funciones, lo que puede facilitar la aparición del sentido.

ETM institucional

Universidad: estudios de ingeniería

El ETM observado en los libros de texto de Budnick (2007) and Haeussler & Paul (2003) no menciona el polinomio de Taylor y el estudio de la optimización se reduce al uso de artefactos simbólicos, es decir, las derivadas, que se utilizan para clasificar los puntos críticos utilizando el criterio de la segunda derivada: el enfoque se centra esencialmente en el paradigma AN2.

En el segundo curso de cálculo universitario, se inicia la optimización de funciones de \mathbf{R}^2 a \mathbf{R} . El trabajo es muy similar al del primer año para las funciones de \mathbf{R} a \mathbf{R} . A pesar de la complejidad de la optimización de funciones de dos variables, los libros de texto trataron de presentarla de la misma manera: en el caso de que el determinante del hessiano sea positivo, se utiliza el signo de $\frac{\partial^2 f}{\partial x^2}(x_0, y_0)$ la segunda derivada parcial con respecto a x , un signo que no es un invariante de la matriz pero que es de gran interés, ya que juega el mismo papel que la segunda derivada para las funciones de una variable (si es positivo es un mínimo, si es negativo es un máximo). Desgraciadamente, no se puede extender fácilmente a más variables. La justificación matemática presentada en ambos manuales no considera la perspectiva local; no hay ningún trabajo en el paradigma AN1. Sin embargo, en uno de los libros de texto se menciona el polinomio de Taylor de segundo orden y se afirma, sin ninguna justificación, que alrededor de un punto la función y el polinomio son similares, pero esto sólo es cierto en el caso en que el determinante de la matriz Hessiana sea distinto de cero.

¿Qué significado tiene para los estudiantes?

Parece que cada vez se le da menos sentido a la optimización de funciones durante la experiencia de aprendizaje de los estudiantes, desde el colegio hasta la universidad. Por ello, pedimos a los estudiantes de ingeniería de segundo año de universidad que resolvieran problemas de optimización. No eran problemas estándar: proporcionamos la superficie que representaba las funciones y, en la mayoría de las tareas, los puntos críticos tenían el determinante de la hessiana igual a cero, pero había un óptimo (bastante) evidente, para $f(x, y) = 2 + x^2 - 2xy + y^2 + x^4$. Esperábamos que los alumnos utilizaran la gráfica o identificaran el óptimo evidente después de intentar aplicar el teorema. Sin embargo, aunque se presenta la gráfica de la función, los alumnos siempre recurren al registro algebraico y a los artefactos simbólicos para utilizar el criterio de las segundas derivadas parciales.

La visualización local

La visualización en el ámbito del análisis matemático (o cálculo) es específica (Kuzniak, Montoya y Vivier, 2016) y la noción de perspectiva local es fundamental. Por supuesto, un trabajo en el paradigma AN2, que abarca lo que comúnmente se llama cálculo, se limita generalmente a perspectivas puntuales y globales. Pero este tipo de trabajo, o bien no permite comprender el trabajo realizado (¿Por qué es creciente cuando f' es positiva?), ya que se apoya esencialmente en artefactos simbólicos (cálculos de derivadas y resolución de inecuaciones, por ejemplo), o bien no se domina y comprende suficientemente para ampliar el abanico de problemas que se pueden resolver (¿Existe un óptimo cuando $f'(a) = f''(a) = 0$?; ¿Cómo se puede extender el teorema de optimización a las funciones de varias variables?; ¿Cómo se pueden tratar los casos degenerados?).

Una de las claves para acceder al significado es resaltar el carácter local de las nociones, que a menudo está contenido en los artefactos simbólicos. El ejemplo de la tangente, una aproximación de orden 1, que se presenta a continuación, ha sido ampliamente tratado en la enseñanza de las matemáticas en este sentido, es decir, mediante el uso de un programa informático que permite ampliar un punto de una curva que representa una función hasta visualizar una línea recta que puede vincularse a la derivada en el punto considerado. La idea es utilizar un artefacto digital para visualizar los signos producidos en la pantalla y así desarrollar el conocimiento matemático, es decir, realizar un trabajo completo. Este es el punto de vista desarrollado posteriormente en nuestra investigación para las aproximaciones de orden 2.

Un experimento con estudiantes de ingeniería

Se diseñó y probó una situación de aprendizaje entre estudiantes de segundo año de ingeniería de la UVM (Universidad Viña del Mar), en un curso llamado “Cálculo en varias variables”. Los estudiantes no tenían conocimientos previos sobre el polinomio de Taylor. Trabajaron en parejas, con GeoGebra, y la discusión entre grupos fue bastante fácil. La atención se centró en dos tareas sobre funciones de una variable: visualizaciones locales y optimización.

El objetivo de la primera tarea era obtener aproximaciones lineales y cuadráticas de una función f en la vecindad de un punto. Se diseñó un trabajo con dimensiones semióticas e instrumentales en el paradigma AN1, utilizando GeoGebra para desarrollar una PL para la aproximación de primer orden (tangente), y luego para la aproximación de segundo orden (parábola osculatriz).

$$\frac{1}{4-x^2}$$

En primer lugar, se pidió a los alumnos que dibujaran la curva de $f(x)=4-x$ en GeoGebra, y la tangente en $A(a, f(a))$ utilizando la herramienta GeoGebra, donde a es un deslizador. GeoGebra crea automáticamente una función afín g para la tangente en la ventana de álgebra. A continuación, se pidió a los alumnos que ampliaran varias veces la imagen y explicaran lo que ocurre gráficamente entre la curva y la tangente en el punto A .

La segunda parte corresponde a una aproximación cuadrática: consistía primero en dibujar la curva de la función $h(x)=g(x)+b(x-a)^2$ donde b es un deslizador. Con los zooms y el deslizador b , se pidió entonces a los alumnos, para tres valores de a en la parte convexa de la curva, -1.6 (curva decreciente), 1.4 y 1.7 (curva creciente), que respondieran a la siguiente pregunta: “¿Cuáles son los valores de b que dan la mejor aproximación local $h(x)$ de la función $f(x)$ alrededor del punto $(1,7,f(1,7))$?” Con esta pregunta, se esperaba que los alumnos obtuvieran los mejores valores de b para los que la parábola se pareciera visualmente a la curva de la función f . La comparación entre $f'(a)$ y b tenía como objetivo rechazar $b=f'(a)$ y conjeturar el valor $b=f'(a)/2$.

La visualización gráfica con aproximaciones de primer y segundo orden a escala local fue una novedad para los alumnos. En general, los resultados obtenidos para el primer orden muestran una PL bastante buena con la visualización de la propiedad de micro-linealidad² (Maschietto, 2002, 2004), a pesar de que algunos estudiantes sólo tomaron una PP inadecuada.

La aproximación cuadrática fue más difícil: la mayoría de los estudiantes adoptaron una PG para responder y, aunque hicieron algunos acercamientos para completar la tabla que se les dio (Fig. 2), lamentablemente no fue suficiente. Es posible rechazar el valor $b=f''(a)$, pero aunque el valor de $f''(a)/2$ se encontraba generalmente en los intervalos dados por los alumnos, estaba cerca del borde de estos intervalos, a veces con una gran longitud, y es bastante difícil hacer una conjetura sobre este valor. Para cada intervalo $[b_1, b_2]$, se calcula $b/f''(a)$ y se compara con 0.5: los valores $0.5 - b_1/f''(a)$ y $b_2/f''(a) - 0.5$, se proporcionan en la Tabla 1. Un valor negativo significa que $f''(a)/2$ no está en el intervalo $[b_1, b_2]$. La conjetura $b=f''(a)$ es bastante clara, incluso con intervalos aproximados, en contraste con la conjetura $b=f''(a)/2$ que necesita valores $|0.5 - b/f''(a)|$ inferiores a 0.1.

Completa la siguiente tabla:

a	-1.6	1.4	1.7
$f''(a)$	7,8232167	2,327536	18,528389
b	3,4 - 5,2	1 - 1,2	7 - 14

Figura 2: Respuesta del Grupo 1

a	-1.6		1.4		1.7	
G2	0.07	0.05	0.07	0.10	0.09	0.01
G1	0.07	0.16	0.07	0.02	0.12	0.26
G7/G8	0.07	0.16	0.16	0.02	0.08	0.00
G4	0.09	0.22	0.20	0.02	0.18	-0.01
G5	0.14	-0.07	0.24	-0.07	0.10	-0.01
G6	0.49	-0.19	0.24	-0.07	0.49	-0.11
G3	0.50	-0.47	0.16	-0.07	0.50	-0.49

Tabla 1: desviaciones $0.5 - b/f''(a)$ y $b/f''(a) - 0.5$ para los intervalos dados por grupos, de los intervalos "mejores" a los "peores"

² Una recta es obtenida entre la función f y la tangente, después de varios zoom con un trabajo instrumental, vemos aparecer una recta, una propiedad de las curvas llamada micro-linealidad en Maschietto 2002, 2004, explotando las ideas de Tall (1985)

En la Tabla 1, podemos ver que sólo G2 tomó una PL completa. Para G1, G7 y G8, la PL fue sólo parcial. Estos grupos parecen haber dado sentido a su trabajo, ya que todos, excepto G7, respondieron “no” a la pregunta “¿Es $b=f''(a)$?”. Sin embargo, sólo G2 conjeturó “ $b=-f''(a)/2$ ”; de hecho, los demás intervalos son demasiado grandes. El intervalo más grande de G1 para $a=1.7$ puede ser el resultado de demasiados zooms en el punto donde la tangente y la curva coinciden en la pantalla.

Al parecer, los grupos G3, G4, G5 y G6 no dieron sentido a su trabajo. A pesar del posible descarte de $b=f''(a)$ por los intervalos dados, ninguno de los otros grupos hizo ninguna conjetura sobre la relación entre b y $f''(a)$, y esperaron a recibir la institucionalización por parte del profesor. Los intervalos encontrados por estos cuatro grupos estaban muy lejos de lo esperado, y muchos de los intervalos no contenían 0.5 (valor negativo en la Tabla 1).

Sin embargo, es bastante difícil entender con precisión el trabajo realizado por los grupos, ya que cada uno posiblemente utilizó diferentes zooms y desplazamientos de pantalla, con diferentes visualizaciones de la relación entre la curva y la parábola.

Según los diferentes puntos a investigar, $a=1.4$ parece más fácil, $a=1.7$ es más difícil, y la visualización es mejor en el lado en el que la pendiente “baja” (a la derecha si $f'(a)<0$ y a la izquierda si $f'(a)>0$). También parece que cuanto mayor sea $|f'(a)|$, peor será la aproximación. Pero podemos esperar una influencia de segundo orden ($f''(a)$ y la curvatura) en la visualización. Estos puntos deberían ser explorados en futuras investigaciones.

La segunda tarea pretende estudiar y clasificar los puntos críticos de $f(x)=\frac{x^4}{2}-2x^3+8$, utilizando el trabajo basado en aproximaciones de Taylor de primer y segundo orden, con dimensiones principalmente semióticas y discursivas en el paradigma AN2. El paradigma AN1 fue favorecido por los estudiantes, que experimentaron muchas dificultades para activar la dimensión discursiva: 5 de los 9 grupos dieron una respuesta basada en lo que observaron gráficamente, y ningún grupo escribió nada sobre los resultados obtenidos algebraicamente. Generalizar sus resultados de forma algebraica fue difícil y los grupos que contestaron lo hicieron de forma incorrecta. Por ejemplo, el grupo más avanzado, G9, escribió: “En el punto mínimo, el polinomio de Taylor forma una parábola cóncava hacia arriba”, sin hacer ninguna conjetura algebraica.

Conclusiones

El resultado más significativo de este estudio se refiere a la visualización local. El fenómeno de la competencia entre PL y PG pone de manifiesto la dificultad que tienen los estudiantes para entender qué es una aproximación local. Estos resultados apoyan los de Montoya Delgadillo et al. (2018) respecto a que la flexibilidad entre perspectivas de localización es importante para el análisis de la enseñanza y el aprendizaje. Nuestro trabajo muestra que el uso de software como GeoGebra puede potencialmente tener un efecto en la visualización, considerando la competencia entre PL y PG. Esta idea innovadora podría utilizarse para enseñar la aproximación de Taylor.

De hecho, los resultados de este estudio sugieren que el trabajo que se da a los estudiantes en la universidad sobre la aproximación de Taylor se realiza principalmente en un paradigma AN2 con el uso de artefactos simbólicos para calcular las derivadas. Por ejemplo, López (2018) analiza el trabajo propuesto a los alumnos de primer curso de matemáticas de la Universidad de París al estudiar el punto crítico de una cardioide: este trabajo se limita al uso del teorema de clasificación de los puntos críticos de las curvas planas parametrizadas, no es justificado ni explicado, utilizando el cálculo de dos aproximaciones de Taylor. Estas aproximaciones de Taylor sólo se utilizaron para determinar los dos enteros que entran en el enunciado del teorema. Sin embargo, estas aproximaciones locales podrían (al menos en este ejemplo) utilizarse para justificar el teorema de clasificación.

Por supuesto, cambiar los hábitos de los profesores universitarios y el trabajo tradicional en el paradigma AN2 sin tecnología sería un proceso a largo plazo. Por ejemplo, en esta investigación, no era habitual que el profesor dejara a los alumnos reflexionar sobre su trabajo y retrasara la institucionalización.

Es difícil saber si el experimento ha logrado su objetivo de permitir a los alumnos dar sentido a la optimización y al estudio local de las funciones. Se carece de datos aunque se haya observado parcialmente el cambio de paradigmas y de representaciones, porque la actividad estaba estrechamente guiada para tener en cuenta el nivel matemático de los alumnos y también por el hecho de que se trataba de un nuevo tipo de trabajo. Es difícil saber si habrían sido capaces de activar las tres dimensiones y articular los planos del ETM solos, en particular la dimensión discursiva en el paradigma AN2. Aunque la visualización gráfica con aproximaciones locales no se logró del todo, los estudiantes lograron trabajar con el software, realizar todas las instrucciones con él como se esperaba y observar los fenómenos generados. Este hallazgo es una contribución para apoyar futuras investigaciones y está en línea con el trabajo de Kidron (2004), quien mostró que en la enseñanza del polinomio de Taylor, las ideas visuales intuitivas ayudan a la comprensión junto con el razonamiento analítico.

Al requerir el uso de la aproximación local, la tarea de optimización fue difícil para los estudiantes. Como mínimo, podemos concluir que hay que dedicar más tiempo a las aproximaciones locales de segundo orden de funciones de una variable (de hecho, se diseñó una nueva tarea con el profesor para trabajar las funciones cuadráticas y las aproximaciones locales). Obviamente, la extensión a funciones de \mathbb{R}^2 a \mathbb{R} fue mucho más difícil para los estudiantes, en particular debido a: la visualización en 3D, la consideración de dos variables en lugar de una, el nuevo trabajo sobre el polinomio de Taylor y la aproximación local, y el hecho de que los estudiantes no tenían conocimientos previos de funciones cuadráticas de dos variables. Todos estos puntos deben ser considerados para futuras investigaciones.

REFERENCIAS BIBLIOGRÁFICAS

- Brousseau, G. 1986. “Fundamentos y métodos de la didáctica”, RDM N° 9 (3). Versión en español publicada por Facultad de Matemática, Astronomía y Física de la Universidad de Córdoba.
- Budnick, F. (2007). *Matemáticas Aplicadas para Administración, Economía y Ciencias Sociales* (Mcgraw-Hill/Interamericana (éd.); Cuarta). Mac Graw Hill.
- Haeussler, E., et Paul, R. (2003). *Cálculo para Administración, Economía, Ciencias Sociales y de la Vida* (A. E. Q. McDonald (éd.); Décima). Pearson Educación.
- López, S. (2018). La algebrización en un curso de curvas y superficies parametrizadas: una sesión de clase. In Montoya, E., Richard, P., Vivier, L., Gómez, I. M., Kuzniak, A., Maschietto, M., and Tanguay, D., editors, *Sixième Symposium sur le Travail Mathématique*, pp. 383–394, Valparaíso, Chili.
- Maschietto, M. (2002). Quelques éléments de l'étude de la transition algèbre analyse au lycée. In *Actes de la 11e école d'été de didactique des mathématiques 21-30 août 2001*. la Pensée Sauvage éditions.
- Maschietto, M. (2004). Le jeu entre point de vue local et point de vue global en analyse : une ingénierie didactique à visée diagnostique au niveau première. In *Actes du colloque de Mulhouse 8-9 mars 2002*. IREM de Strasbourg.
- Montoya, E., et Vivier, L. (2016). Mathematical working space and paradigms as an analysis tool for the teaching and learning of analysis. *ZDM - Mathematics Education*, 48(6), 739-754. <https://doi.org/10.1007/s11858-016-0777-9>.
- Montoya Delgadillo, E., Páez Murrillo, R., Vandebrouck, F., & Vivier, L. (2018). Deconstruction with Localization Perspective in the Learning of Analysis. *International Journal of Research in Undergraduate Mathematics Education*, 4(1), pp. 139-160.
- Kuzniak, A., Tanguay, D., & Elia, I. (2016). Mathematical Working Spaces in schooling: an introduction. *ZDM*, 48, pp. 721–737.
- Vandebrouck, F. (2011). Perspectives et domaines de travail pour l'étude des fonctions. *Annales de Didactiques et de Sciences Cognitives*, 16, pp. 149–185.

2

Un acercamiento a la Neuroeducación

Patricia Cabrera – Calavacero

*Escuela de Ciencias
pcabrera@uvm.cl*

RESUMEN

Este trabajo se plantea como una breve revisión del nacimiento de la Neurociencia y de cómo esta ha entregado información valiosa a través de diferentes investigaciones sobre el sistema nervioso y en particular cómo funciona el cerebro en el proceso de aprendizaje. En primer lugar, se mencionará el origen de una nueva disciplina llamada Neurociencia Educativa o Neuroeducación, la cual está en proceso de búsqueda de respuestas relacionadas a las estrategias de aprendizaje. Seguidamente se analizará la relación establecida entre aprendizaje y memoria, aprendizaje y motivación, entre otras y se mencionarán los aportes que ha entregado la Neurociencia educativa con relación a los neuromitos y metodologías innovadoras tales como la gamificación. Finalmente, se mencionarán algunos aportes de la Neurociencia a la investigación como un nuevo paradigma.

Palabras clave: Neuronas, cognición, aprendizaje, Neuroeducación.

Introducción

Una pregunta que muchos docentes nos hacemos en el inicio de una asignatura es ¿Cómo poder lograr que mis alumnos aprendan? Las teorías que se han establecido sobre el aprendizaje difieren mucho sobre cómo se aprende y existe una discrepancia en torno a la función que asignan a los procesos motivacionales. Las teorías conductuales no distinguen motivación de aprendizaje, sino que se sirven de los mismos principios para explicar cualquier conducta. Sin embargo, casi todas las teorías contemporáneas opinan que motivación y aprendizaje se relacionan, pero no son sinónimos (Brophy,1983).

Definimos la cognición como un conjunto de procesos que permiten el procesamiento de la información y el desarrollo del conocimiento. (OCDE,2009). Dichos procesos se denominan “funciones cognitivas” y son las que nos permiten llevar a cabo cualquier tarea.

Entre las funciones cognitivas podemos mencionar la atención, la orientación, la memoria, las funciones ejecutivas, el razonamiento, el lenguaje y la toma de decisiones, entre otras.

El inicio de la Neuroeducación

El estudio del cerebro se remonta a la prehistoria, donde se ha encontrado evidencia de la realización de trepanaciones, que consistía en una práctica médica en la que se agujereaba el cráneo.

A finales del siglo XVIII se descubrió la actividad eléctrica del sistema nervioso, lo que dio origen a los análisis en el campo de la electrofisiología neuronal. Un siglo más tarde se descubrió el mecanismo mediante el cual se comunican las neuronas, llamado transmisión sináptica.

En los años 60 nace la Neurociencia como un estudio disciplinar gracias a los avances tecnológicos de la época. Se ha reconocido a Ramón y Cajal como el padre de la Neurociencia por su identificación de la neurona como unidad básica, pero datos asociados a fechas indican que sería más adecuado establecer a Sherrington como pionero de esta ciencia.

La Neurociencia nace como un paraguas epistemológico que reúne a diferentes disciplinas (física, psicología, filosofía, medicina, biología, química, entre otras) con la intención de conocer la estructura, la función, el desarrollo, la bioquímica, el funcionamiento neuronal y la patología del sistema nervioso, así como la forma en que sus diferentes elementos interactúan, dando lugar a las bases biológicas de la conducta (Blakemore, Frith,2011)

Dentro de las aplicaciones de esta disciplina se encuentran los aportes al área de la educación, a través de la neurociencia educativa o neuroeducación, disciplina que fusiona los conocimientos de la neurociencia, psicología y educación. Carballo (2018) la define como “una disciplina que pretende integrar los conocimientos neurocientíficos acerca de cómo funciona y aprende el cerebro en el ámbito educativo”.

El fin de la neuroeducación se podría decir que es producir una mejora en los métodos de enseñanza que se están utilizando en la actualidad, ya que se ha visto cómo la forma de realizar clases requiere en forma urgente de innovación.

Aprendizaje y memoria

Según la definición establecida por Purves et al. en el 2008, aprendizaje es el nombre dado al proceso por el que el sistema nervioso adquiere nueva información que se observa mediante cambios en el comportamiento. Por otro lado, la memoria se refiere a la codificación, el almacenamiento y recuperación de la información aprendida.

Dentro de la clasificación cualitativa de la memoria de largo plazo podemos establecer dos sistemas de almacenamiento que son la memoria declarativa, que es la que contiene información de la cual se tiene un registro consciente y que se puede verbalizar, y la memoria no declarativa que comprende información que no se puede verbalizar fácilmente o cuyo aprendizaje se realiza en forma involuntaria. Además de esa memoria también se ha podido establecer que existe la memoria de corto plazo donde podemos diferenciar la inmediata (fracciones de segundos) y la memoria de trabajo (segundo-minutos).

Se sabe que la capacidad para recordar la información es limitada, podemos recordar entre 7 a 9 números o elementos arbitrarios. Dicha capacidad se puede aumentar utilizando estrategias que permitan entregar un significado a algo que no lo tiene.

Existe el proceso de consolidación de la memoria donde a través de un proceso gradual de la reiteración de la memoria de corto plazo se originan cambios neuronales que dan paso a la memoria a largo plazo. Respecto a esto, el psicólogo alemán Ebbinghaus (1850-1909) fue uno de los primeros autores que estudiaron la memoria. Ebbinghaus aceptaba los principios de asociación y creía que el aprendizaje y el recuerdo de la información aprendida depende de la frecuencia de la exposición al material (Schunk, 1997).

Una forma de demostrar dicha hipótesis fue realizar un experimento en el que él participó aprendiendo sílabas sin sentido de tres letras, las cuales visualizaba una a una mirándolas detenidamente por un lapsus de tiempo. Observó que mientras más revisaba la lista cometía menos errores. Dicho experimento lo realizó también con fragmentos con sentido, obteniendo la conclusión de que se hace más fácil aprender aquello que sea significativo.

Sus estudios dieron origen a la llamada curva del olvido que es una función logarítmica que muestra que tras repasos espaciados en el tiempo, la curva se va suavizando, mostrando que la pérdida de memoria después de cada repaso es menor.

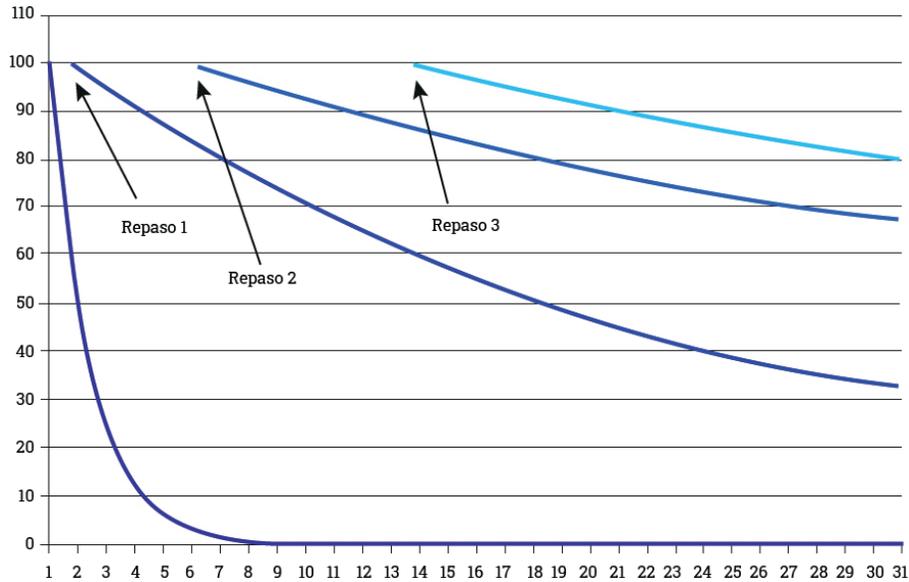


Figura 1: Curva del olvido comparada con las curvas con repaso regulares (imagen extraída de <https://neurodidactic.com>)

El Aprendizaje y las emociones

Las emociones son reguladas por el sistema límbico o también llamado “centro de las emociones”, que está conformado principalmente por la amígdala y el hipocampo, además de estructuras corticales que se encuentran principalmente en la corteza prefrontal.

Cuando nuestro cerebro participa en una experiencia estimulante, más de cien mil millones de neuronas se conectan entre ellas. La dopamina es uno de los neurotransmisores que utilizan las neuronas para comunicarse y el aumento de sus niveles genera emociones positivas y agradables.

La relación aprendizaje y emoción hace algún tiempo ha estado en la palestra ya que muchos pensaban “que las emociones eran secundarias a la cognición en el aprendizaje, pero investigaciones Neurocientíficas y del desarrollo revelan que ambas están entrelazadas” (immordino – Yang, Damasio., 2007). Las emociones no se pueden considerar separadas del proceso cognitivo, ya que estas facilitan el pensamiento racional apoyando la toma de decisiones a través de una retroalimentación emocional.

En la investigación realizada por Morrens y sus colegas, donde se monitorearon las señales de las neuronas dopaminérgicas en el área ventral tegmental y la sustancia nigra

de ratones, se encontró que al proporcionar estímulos nuevos el aprendizaje se produce más rápido.

La alegría se asocia con altos niveles de dopamina en el sistema de recompensas del cerebro. Sistema relacionado con la ampliación de la memoria, la atención, cambios mentales, la creatividad y la emoción.

Aplicación de la Neurociencia en el aprendizaje

Dentro de los aportes más importantes de la Neurociencia a la educación están las estrategias educativas y terapéuticas que generó en torno a los trastornos de aprendizaje y de desarrollo. Se han realizado intervenciones concretas asociadas a las áreas específicas asociadas a los trastornos como dislexia, discalculia, disgrafía, déficit atencional y trastornos del espectro autista.

Un tema relevante es el uso de creencias pseudocientíficas en educación, las cuales muchos docentes han llevado a sus prácticas. Entre ellas podemos mencionar los Neuromitos, definidos como falsas creencias que se tiene sobre el desarrollo cerebral. En un texto escrito por la OCDE (2007) se menciona algunas tales como:

1. “No hay tiempo que perder ya que todo lo importante para el cerebro está decidido los tres años de edad”
2. “Existen períodos críticos cuando se debe enseñar y aprender ciertas materias”
3. “Sin embargo he leído en alguna parte que usamos solamente 10% de nuestro cerebro”
4. “Soy una persona del ‘hemisferio izquierdo’, ella es una persona del ‘hemisferio derecho’
5. “Aceptémoslo: los hombres y los niños tienen cerebros diferentes de los de las mujeres y las niñas
6. “El cerebro de un niño pequeño solo puede manejar el aprendizaje de un idioma a la vez”
7. ¡Mejore su memoria!
8. Aprenda mientras duerme

En los últimos años la gamificación, es decir, el aprendizaje basado en juego ha sido una de las metodologías que se ha utilizado para implementar en las aulas. Su carácter lúdico facilita la interiorización de los conocimientos de una forma más divertida, generando una experiencia positiva en quien la utiliza.

La gamificación social consigue implicar a los alumnos a través de la emoción, activando la amígdala que es una estructura subcortical que según la Neuroeducación provoca el aprendizaje emocional, el aprendizaje asociativo y el aprendizaje condicionado.

Los estudiantes que actualmente están ingresando a la universidad, son jóvenes nacidos, criados y educados en ambientes cargados de tecnología y que esperan que el aprendizaje sea lo más rápido, sencillo y entretenido posible, maximizando la relación entre resultados obtenidos y tiempo de estudio. (Ibáñez, Cuesta, Taglibaue & Zangaro, 2008).

Las herramientas tales como el Kahoot, Quizizz, entre otros se han utilizado como caballos de batalla para la implementación de dicha metodología generando nuevos desafíos en los docentes, quienes deben en la actualidad potenciar habilidades relacionadas con el manejo de la tecnología.

Aportes de la Neurociencias a la investigación

La investigación en general en educación ha respondido a dos tipos de paradigmas: el positivista, que tiene como objetivo explicar, relacionar y predecir variables y el interpretativo, al que le interesa fundamentalmente la comprensión de los fenómenos, de los procesos, los hechos.

En el caso del paradigma positivista, las técnicas a utilizar son instrumentos tipo test, cuestionarios y observación sistemática entre otros. En el caso del paradigma interpretativo se utilizan entrevistas, observaciones, participantes, análisis de documentos, diarios, etc.

En la época de los 70 se observó que tanto los métodos cualitativos y cuantitativos, si bien permitían explicar muchos procesos, fenómenos o hechos, no permitían mejorarlos ni transformarlos. Esto dio origen al paradigma crítico cuyo propósito de conocimiento está centrado en el cambio y la transformación de la realidad educativa o social en los diferentes contextos. En este caso las técnicas utilizadas abarcan la investigación - acción, investigación evaluativa y estudio de diseño.

La mayoría de los autores está de acuerdo con que un proyecto de investigación basado en diseño debe desarrollarse a través de ciclos de diseño, implementación, análisis y rediseño.

REFERENCIAS BIBLIOGRÁFICAS

- Blakemore, S. J., Frith, U. (2011). *Cómo aprende el cerebro. Las claves para la Educación.* Ariel: Barcelona
- Brophy, J. (1983). Investigación sobre la profecía autocumplida y las expectativas del maestro. *Revista de Psicología de la Educación*, 76, 236-247.
- Carballo, Anna., Portero, M. (2018). *10 ideas claves. Neurociencia y Educación (1a.ed.)* Barcelona. Editorial GRAÓ, de IRIF, S.L.
- Ibáñez, E., Cuesta, M., Taglibaue, R & Zangaro, M. (2008). La generación actual en la universidad: el impacto de los millennials. V Jornadas de Sociología de la UNLP. Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Departamento de Sociología. La Plata. Recuperado de <https://www.aacademica.org/000-096/261.pd>
- Immordino-Yang, Mary & Damasio, Antonio. (2007). *We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. Mind, Brain, and Education.* 1. 3 - 10. 10.1111/j.1751-228X.2007.00004. x.
- Liu, C., Solis, S.L, Jensen, H., Hopkins, E. J., Neale, D., Zosh, J. M., Hirsh-Pasek K, & Whitebread. D. (2017) *La Neurociencia y el aprendizaje a través del juego: un resumen de la evidencia (resumen de investigación).* The LEGO Foundation, DK
- Morrens, J., Aydin, C., Janse van Rensburg, A., Esquivelzeta Rabell, J., & Haesler, S. (2020). Cue - evoked dopamine promotes conditioned responding during learning. *Neuron*,106(1),142-153
- OCDE. (2009). *La comprensión del cerebro: el nacimiento de una ciencia del aprendizaje.* Ediciones Universidad Católica Silva Henríquez.
- Purves, D., Augustine, G.j., Fitzpatrick, D., Hall, William. C., Lamantia, A.S., McNamara, J.O., Williams, S.M. (2008) *Neurociencia (3ra edición).* Editorial medica Panamericana.
- Schunk, D.H (1997). *Teorías del aprendizaje (Vol.2).* Prentice -Hall, Inc. A. Simon & Schuster Company.

3

La Media Derivada

Neumarino Rodríguez

*Escuela de Ciencias
nrodriguez@uvm.cl*

RESUMEN

La media derivada corresponde a lo que se conoce como "Cálculo Fraccionario". Este tema de la matemática es relevante en la simulación de una serie de fenómenos en el área de las matemáticas aplicadas. En este ensayo se pretende presentar una conceptualización operativa para su comprensión operativa. Lo anterior se realiza mediante un estudio socio epistemológico. La propuesta técnica utiliza una aproximación a través de la función gamma. Actualmente es un tema de alta investigación en matemática y que incide en la simulación de una serie de fenómenos.

Palabras Claves: media derivada, calculo fraccionario, matemáticos

Introducción

Es bien conocida la controversia iniciada en el siglo XVII entre I. Newton (1678-1778) y G. Leibnitz (1646-1716) respecto de la invención del Cálculo. Estos científicos resolvieron el mismo problema desde focos diferentes. En este ensayo iniciaremos la partida desde una publicación realizada por Leibnitz en el cual se introduce la noción de derivada y se utiliza una notación que ha perdurado hasta nuestros días, a saber, $\frac{d^n y}{dx^n}$ donde $n \in \mathbb{N}$.

La publicación por parte de Leibnitz llegó a la mano de G. Hôpital (1661-1704), quien escribe una carta a Leibnitz en septiembre de 1697 donde pregunta ¿Qué sucedería si n fuera $\frac{1}{2}$?

Responde Leibnitz:

“...esto conduciría a una paradoja, de la cual algún día se extraerán consecuencias útiles...”

Un Socioepistemología Inversa

Iniciaremos una mirada desde la notación usual para la derivada de orden superior de una función. Consideremos la función, $f(x) = x^m$, con derivada de orden n , $n \in \mathbb{N}$.

Es conocido que:

$$\frac{d^1 y}{dx^1} = mx^{m-1}; \quad \frac{d^2 y}{dx^2} = m(m-1)x^{m-2}; \quad \frac{d^3 y}{dx^3} = m(m-1)(m-3)x^{m-3}; \dots$$

Siguiendo con la estructura anterior se concluye que:

$$\frac{d^n y}{dx^n} = \frac{m!}{(m-n)!} x^{m-n}, \quad \text{con } m \geq n.$$

En 1819 Lacroix (1765-1843) utilizando el símbolo Γ , introducido por A. Legendre para la función gamma, obtuvo la siguiente expresión

$$\frac{d^n y}{dx^n} = \frac{\Gamma(m+1)}{\Gamma(m-n+1)} x^{m-n}$$

La función gamma, anotada por $\Gamma(z)$ con $z \in \mathbb{C}$ es una función que extiende el concepto de factorial a los números Reales y Complejos, y se define por:

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt \quad \text{con } z \in \mathbb{C} \text{ y } \text{Re}(z) > 0$$

Si $n \in \mathbb{N}$ entonces $\Gamma(n) = (n-1)!$, lo que muestra la relación entre la función gamma y la factorial.

Es así, como Lacroix anota simplemente como respuesta a la pregunta de L'Hôpital en el caso de $f(x) = \sqrt{x}$ la expresión:

$$\frac{d^{\frac{1}{2}} y}{dx^{\frac{1}{2}}} = \frac{\Gamma(2)}{\Gamma(\frac{3}{2})} x^{\frac{1}{2}} = \frac{1! x^{\frac{1}{2}}}{\frac{1}{2} \Gamma(\frac{1}{2})} = \frac{2\sqrt{x}}{\sqrt{\pi}}$$

El resultado anterior ha permanecido como respuesta inalterable hasta hoy, es decir, ha resistido los distintos avances en la formalidad que el cálculo fraccionario ha dado respuesta a la media derivada.

Utilizando lo establecido por Lacroix, anotamos algunos resultados adicionales para potencias superiores a uno para la función potencial $f(x) = x^m$

$$m = 2 \quad y \quad n = \frac{1}{2} \Rightarrow \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(x^2) = \frac{\Gamma(3)}{\Gamma(\frac{5}{2})} x^{\frac{3}{2}}, \text{ por lo tanto } \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(x^2) = \frac{8\sqrt{x^3}}{3\sqrt{\pi}}$$

$$m = 3 \quad y \quad n = \frac{1}{2} \Rightarrow \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(x^3) = \frac{\Gamma(4)}{\Gamma(2)} x^{\frac{5}{2}}, \text{ por lo tanto } \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(x^2) = \frac{120\sqrt{x^5}}{6\sqrt{\pi}}$$

$$m = 0 \quad y \quad n = \frac{1}{2} \Rightarrow \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(x^0) = \frac{\Gamma(1)}{\Gamma(\frac{1}{2})} x^{\frac{1}{2}}, \text{ por lo tanto } \frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}}(1) = \frac{1}{\sqrt{\pi x}}$$

Lo expuesto anteriormente resuelve la pregunta para un tipo muy restringido de funciones y, en consecuencia, persistió la pregunta por muchos años. En 1822 Fourier (1768-1830) retoma el camino hacia el cálculo fraccionario, pero al igual que otros matemáticos de renombre, a saber: Euler, Laplace y otros no hubo mayor aporte a la formalización. En 1823 Abel (1802-1829), matemático noruego quien falleció a los 26 años a causa de la tuberculosis, aplicó por primera vez el cálculo fraccionario en la solución de un conocido problema de la época, denominado, “El problema de la tautócra”.

Posteriormente han existido abundantes discusiones de matemáticos renombrados, algunos seguidores de Lacroix y otros tomando partido por la aproximación planteada por Liouville; También hubo miradas para integrar ambas conceptualizaciones como las de Morgan (1806-1871), matemático británico.

Actualmente se considera a Mittag-Leffler (1846-1927), matemático sueco, y Marcel (1886-1969), matemático húngaro, como aquellos que dieron origen a una caracterización consistente del cálculo fraccionario.

Aplicaciones

Maravall (1923-2016) Ingeniero agrónomo español es reconocido como el primero en mencionar unas oscilaciones fraccionarias relacionadas con ecuaciones diferenciales no enteras.

Caputo (1927-), matemático italiano, estableció la definición de derivada fraccionaria como:

$$\frac{d^\alpha f(x)}{dx^\alpha} = \frac{1}{\Gamma(\alpha-n)} \int_a^x \frac{f^{(n)}(u)}{(x-u)^{\alpha-n+1}} du, \text{ con } \alpha \in \mathbb{R}, n \in \mathbb{N}, \text{ tal que } n-1 < \alpha < n$$

REFERENCIAS BIBLIOGRÁFICAS

- Dorrego, G., Luque., & Cerutti, R. (2021). Cálculo Fraccionario y K-Funciones. Argentina: 978-987-4029-13-3.
- Guía, M., Rosales, J., Guzmán, R., González, A. (2015). El cálculo diferencial e integral fraccionario y sus aplicaciones. mayo 21,2020, de Acta Universitaria Sitio web: repositorio.ugto.mx
- Vázquez, M.&Velasco, M. (2012). Cálculo Fraccionario Como Instrumento de Modelización. España: Académica Española.

4

Presión de radiación y sombras en discos protoplanetarios

Matías Montesinos

Escuela de Ciencias
matias.montesinos@uvm.cl

RESUMEN

Los planetas se forman al interior de discos de gas y polvo que rotan alrededor de estrellas jóvenes llamados discos protoplanetarios. Hasta la fecha, ha sido extremadamente difícil observar planetas durante el proceso de formación. Recientes observaciones de ALMA/SPHERE han revelado evidencia directa e indirecta de planetas embebidos en estos discos. Las características más notorias son; fuentes puntuales luminosas, anillos en el gas y polvo, cavidades, y sombras proyectadas en el disco, interpretadas como interacciones entre planetas y el disco madre.

La interpretación de estas observaciones requiere un robusto escenario físico para deducir esta compleja geometría del disco. Normalmente, los modelos de discos protoplanetarios asumen que su altura viene dada por la presión del gas del disco, ignorando, por ejemplo, las fuerzas provenientes de radiaciones internas del disco i.e., presión de radiación. Despreciar estos términos podría llevar a interpretaciones incorrectas sobre la real altura de los discos observados.

En este trabajo, mediante simulaciones hidrodinámicas, modelamos la evolución de un disco protoplanetario con un planeta tipo Júpiter en su interior, calculando la altura del disco incluyendo términos de presión radiativa, derivada a partir de modelos de atmósfera estelar, poniendo especial atención en el impacto que esta presión radiativa tiene en las zonas de un planeta en formación, donde las temperaturas alcanzan valores $> 1000\text{K}$, y las fuerzas radiativas pueden sobrepasar las fuerzas gravitacionales del planeta.

Encontramos que estas fuerzas de radiación crean una columna vertical de gas sobre la región del planeta. Al ser ópticamente opaca y sobrepasar la altura normal del disco, esta columna proyecta una sombra en las regiones externas al disco, la cual se mueve a velocidad kepleriana junto al planeta y es potencialmente observable en luz scattering ($\sim \mu\text{m}$). Recientes observaciones usando VLT del disco al rededor de HD169142 sugieren la presencia de sombras con características similares a las predichas teóricamente en este trabajo.

Palabras clave: discos protoplanetarios – simulaciones hidrodinámicas – transferencia radiativa – formación de planetas

Introducción

Durante los últimos años, imágenes en alta resolución angular de discos alrededor de estrellas jóvenes han revelado diferentes patrones en la estructura de gas y polvo que componen estos discos. Estas estructuras son normalmente observadas en los discos de estrellas de masa intermedia o baja, presentan una cavidad interior y son llamados discos de transición. Estas imágenes han sido espectaculares por la cantidad de detalles que se han podido observar. Por ejemplo, en el disco al rededor de HL tau, ALMA reveló, por vez primera, una serie de anillos y cavidades, observables en el milimétrico (polvo) (ALMA Partnership et al., 2015). Existen una serie de otros ejemplos notables de observaciones en alta resolución tanto en el infrarrojo como el sub-milimétrico. Se piensa que en gran medida estas estructuras son ocasionadas por planetas en formación al interior de estos discos protoplanetarios (e.g., Long et al. 2018). Un resumen completo de estas observaciones puede ser encontrado en la serie de artículos del proyecto DSHARP (Adrews et al., 2018; Huang et al., 2018a,b).

Aparte de estas impresionantes morfologías observadas en el gas y polvo, intrigantes patrones en la iluminación del disco han sido detectados. Por ejemplo, Avenhaus et al. (2014) reportó dos sombras observadas en el disco en HD142527, las cuales fueron interpretadas más adelante como resultados de un mini disco inclinado rodeando a la estrella, ocultando de esta forma una fracción de la radiación estelar y proyectando dos sombras (una por cada extremo del mini disco) al disco circumestelar exterior (Marino et al., 2015). En estos últimos años se han encontrado, usando polarimetría diferencial con VLT/SPHERE, nuevas evidencias de sombras proyectadas en otros sistemas (Stolker et al., 2016; Debes et al., 2017).

De manera muy interesante, Quanz et al. (2013) reportó observaciones, usando VLT/NACO, del disco alrededor de HD169142, mostrando una delgada sombra localizada a ~ 80 grados. Recientemente, Bertrang et al., (2018, 2020) presentó datos VLT/SPHERE del mismo sistema, pero a mayor resolución angular, revelando una sombra pero que ahora se encuentra ubicada a ~ 50 grados, sugiriendo que se trata de la misma sombra que se ha movido.

Muchos de estos efectos pueden ser atribuidos a la presencia de planetas en formación que al interactuar con el disco madre van moldeando la estructura de su densidad. Sin embargo, la detección directa de planetas en formación ha sido elusiva. Hasta la fecha, sólo un planeta en formación ha sido detectado de forma directa : es el caso del planeta en la cavidad observada en el disco en PDS70 (Keppler et al., 2018; Müller et al., 2018).

Una cantidad importante para describir la geometría de un disco es su escala de altura H/r , dónde H es la altura en un punto del plano medio y r la distancia desde la estrella a este punto. Esta cantidad es necesaria para describir muchas observaciones, especialmente aquellas no resueltas. Por ejemplo, los discos de transición exhiben un exceso en emisiones infrarrojo cercano (NIR), las cuales pueden ser explicadas considerando altos valores de $H/r \sim 0.2$ en el borde interno del disco (Natta et al., 2001; Dullemond et al., 2001; Olofsson et al., 2013).

La extensión vertical de un disco está relacionada con la gravedad (principalmente de la estrella) y con la presión de gas del disco. En general, en los discos circumstelares se espera que domine la componente presión de gas, la cual es equilibrada por la gravedad de la estrella, dando origen a la relación $H/r = cs / vk$, dónde cs y vk son la velocidad del sonido, y la velocidad Kepleriana del gas, respectivamente.

Sin embargo, vale la pena preguntarse si esta condición se mantiene en regiones del disco que se encuentre a alta temperatura ($>1000K$), en donde fuerzas radiativas podrían ser dominantes, como por ejemplo en las inmediaciones de un planeta en formación, o regiones muy cerca de la estrella. Durante la formación planetaria, los protoplanetas acretao son procesos con luminosidades intrínsecas del orden $10^{-6}-10^{-3}L_{\text{sun}}$, y temperaturas de $1000-2000K$ (Montesinos et al., 2015).

A altas temperaturas, la presión radiativa es no-despreciable, similar a lo que pasa en los procesos estelares (Mihalas & Mihalas 1984), donde la escala de altura se ve dominada por fuerzas radiativas (Montesinos & de Freitas Pacheco, 2011).

En este trabajo se adoptó una generalización del modelo de atmósferas estelares aplicadas a la geometría de discos introducida por Hubeny (1990) para considerar presiones de radiación a modo de calcular la escala de altura H/r de un disco protoplanetario. Sumado a esto, se consideró la gravedad del planeta para obtener la altura final.

Encontramos que alrededor de planetas acretao calientes, las fuerzas radiativas dominan sobre la gravedad del planeta, levantando una columna de gas ópticamente opaca en el disco circumplantario que orbita alrededor del planeta. Esta columna de gas bloquea cierta fracción de la radiación estelar, proyectando una sombra en la superficie del exterior del disco. Esta columna puede explicar las sombras observadas en HD161942, las cuales se mueven a velocidad Kepleriana (Bertrang et al., 2018, 2020). Su origen sería entonces la presencia de un planeta aún no detectado orbitando cerca de su estrella

Desarrollo

Mediante simulaciones numéricas usando el código FARGO (Masster 2002), se modeló la evolución de un disco protoplanetario con un planeta tipo Júpiter, ubicado a 10 au de la estrella, considerando una luminosidad planetaria intrínseca de $10^{-3}L_{\text{sun}}$. A esta luminosidad se le llama feedback del planeta. Cuando el feedback del planeta es activado en la simulación, la región alrededor del planeta alcanza temperaturas del orden $\sim 1060\text{K}$ en nuestro modelo. Estas luminosidades se supone deberían estar presentes en las fases temprana de un planeta en formación, la cual podría durar $\sim 1\text{Myr}$ (Mordaisni et al., 2012).

La estructura vertical del disco es descrita tomando en cuenta la presión del gas del fluido, junto a la presión de radiación de este. El gradiente de la presión resultante es equilibrado con la gravedad de la estrella y del planeta siguiendo el procedimiento de Müller et al., 2012.

El tratamiento para la presión de radiación proviene de una prescripción analítica usada en la teoría clásica de atmósferas estelares, adaptada a la geometría de un disco (Hubeny 1990).

Nuestros resultados muestran que, en equilibrio termodinámico, una presión radiativa no despreciable aparece en las cercanías de un planeta acretando cuando su temperatura supera los 1000K . Esta presión de radiación sobrepasa la gravedad del planeta, levantando una columna de gas opaco en la dirección vertical del planeta. Cuando el planeta no está acretando, su luminosidad es $L_p = 0$ y no se observa esta columna de gas.

En la Figura 1, se aprecia la columna de gas que se genera sobre el planeta. Esta aparece solamente al incluir presión de radiación en nuestros cálculos. La altura de esta columna alcanza aproximadamente 2 au, para un planeta localizado a 10 au de la estrella. Esto da una escala de altura de $H/r \sim 0.2$. En el caso contrario, con $L_p = 0$, el gas en la dirección vertical del planeta colapsa debido a la gravedad de este (Fig.1, arriba).

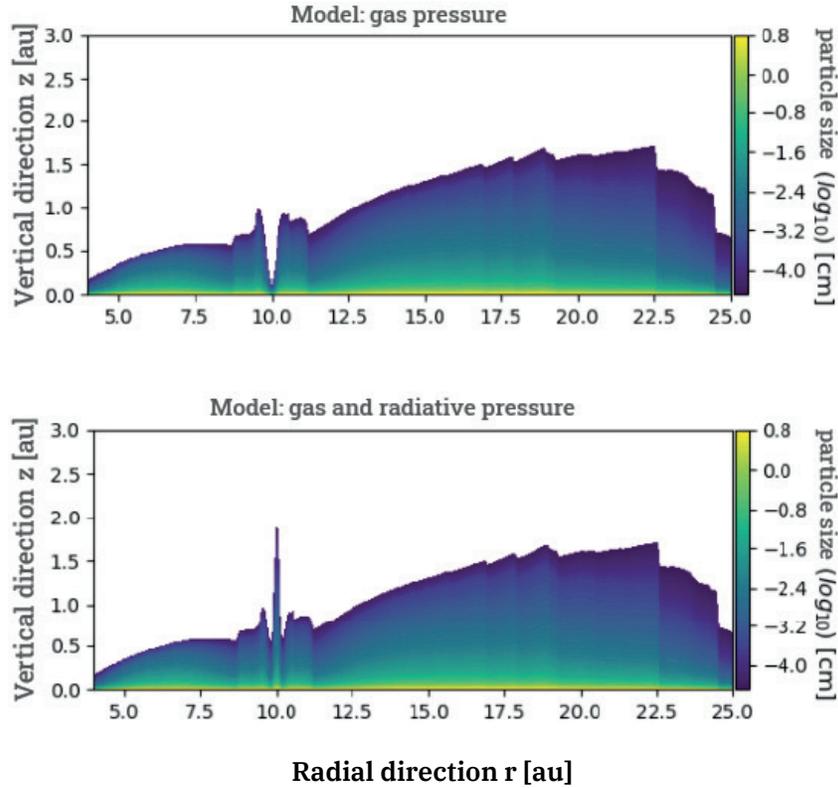


Figura 1: Distribución de tamaños de polvo del disco en el plano r, z . El planeta, con feedback $L_p = 10^8 L_{sun}$. Está ubicado a $r = 10$ au. La parte superior corresponde a un modelo sin presión de radiación, mientras que en la inferior se han incluido los efectos radiativos. Para este caso, se observa como la presión de radiación es capaz de levantar una columna de polvo por sobre la vertical.

Luego de un cálculo de transferencia radiativa, usando el código RADMC3D (Dullemond et al., 2012), podemos obtener una imagen sintética del disco. En la Figura 2, se comparan los modelos de transferencia radiativa con y sin feedback, incluyendo presión de radiación o sólo presión del gas. En esta imagen podemos observar una sombra que aparece proyectada desde el planeta hacia el exterior del disco, la cual solamente se produce si y sólo si: i) se ha incluido presión de radiación, ii) el planeta tiene feedback activado.

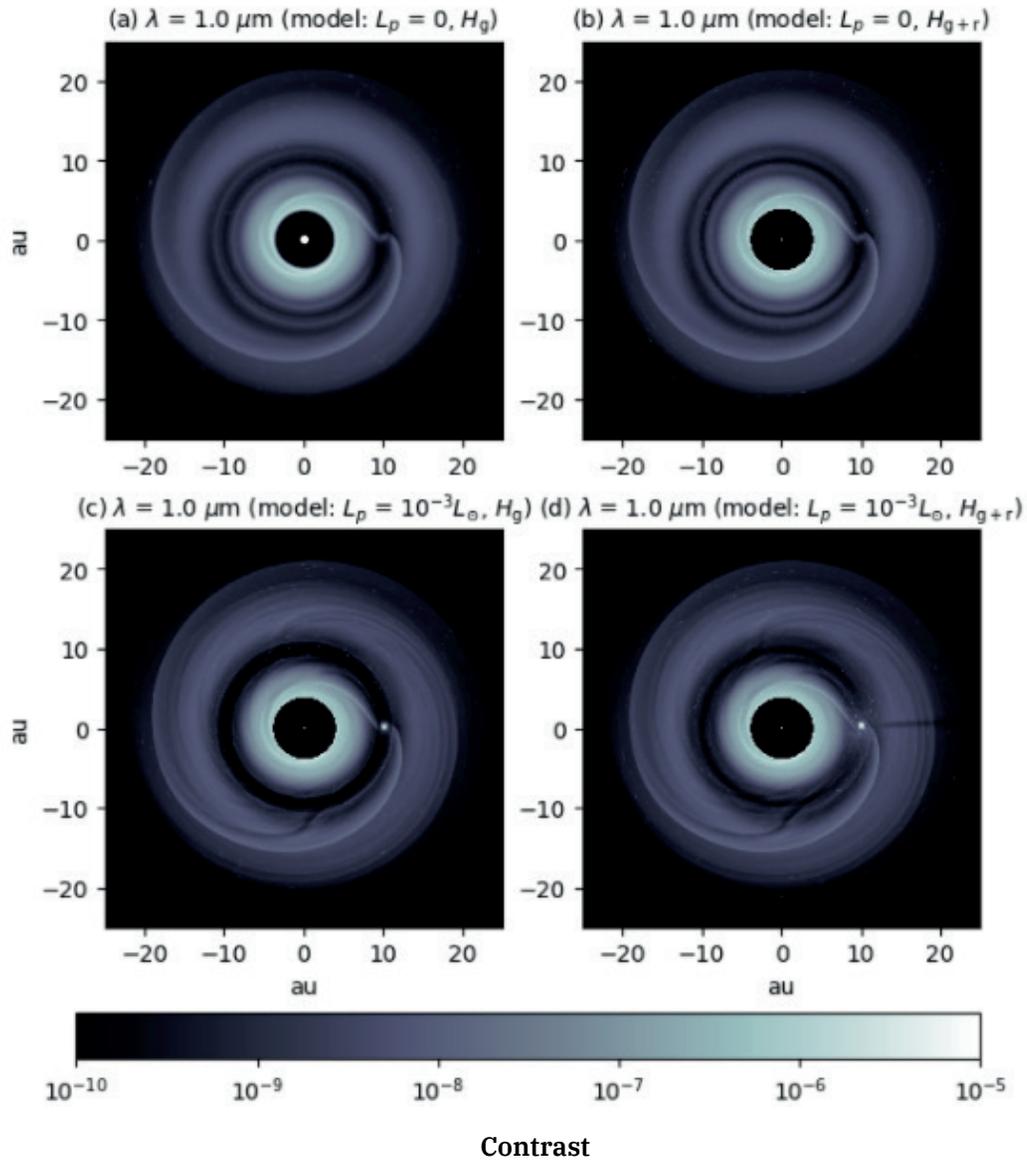


Figura 2: Imágenes sintéticas para $\lambda = 1 \mu\text{m}$ obtenidas mediante RADMC3D. Panel (a): Modelo sin feedback ($L_p=0$), donde sólo se consideró la presión de gas. Panel (b) Mismo modelo sin feedback, pero incluyendo presión de radiación. Panel (c): Modelo con feedback ($L_p=10^{-3}L_{\text{sun}}$) sin presión de radiación. Panel (d): Modelo con feedback y presión de radiación.

La imagen mostrando la sombra proyectada desde el planeta (Fig. 2), demuestra que la inclusión del término de presión radiativa genera una columna de gas opaca sobre el planeta, la cual sólo es observable a longitudes de onda de $\sim 1 \mu\text{m}$, esto debido a que sólo las partículas pequeñas de ese tamaño son levantadas por la presión de radiación, haciendo esta columna de gas opaca para esas frecuencias. Para partículas más grandes, estas se asientan en el plano medio del disco, luego la sombra desaparece para longitudes de onda mayores (milimétricas).

Finalmente, reconciliemos nuestros cálculos teóricos con observaciones recientes obtenidas con ZIMPOL/SPHERE por Bertrang et al., (2018, 2020), donde se revela una sombra que ha cambiado su posición desde su primera observación en 2013 (Quanz et al., 2013). De acuerdo a Bertrang et al., (2020), esta sombra se mueve a razón de 10 grados por año.

De acuerdo a nuestro modelo, esta sombra observada sería producida por un planeta en formación que aún no ha sido observado, el cual estaría localizado a ~ 13 au de la estrella.

Un caso interesante a analizar en el futuro, corresponde a la observación directa de PDS70b (Keppler et al., 2018; Müller et al., 2018), donde se ha sugerido que corresponde a un planeta acreando de $10M_{\text{J}}$. Bajo estas circunstancias, una sombra debería ser proyectada en la región externa del disco al rededor de PDS70. De hecho, notablemente se han encontrado ciertas evidencias de sombras en las regiones externas (Long et al., 2018b), sin embargo su origen es aún incierto. Observaciones en alta resolución del polco micrométrico en PDS70 podría, en principio, revelar nuevas sombras originadas en las regiones cercanas a PDS70b. Este sistema es actualmente un excelente laboratorio para testear nuestras ideas.

Conclusiones

En este trabajo mostramos que en un disco protoplanetario, cualquier región con temperaturas $> 1000\text{K}$, deben incluirse presiones de radiación a fin de calcular la escala de altura del disco. La presión de radiación puede describirse mediante la expresión analítica presentada por Hubeney (1990) o equivalentemente, las ecuaciones presentadas en Montesinos & de Freitas Pacheco (2010).

En regiones cercanas a un planeta acreando, donde se esperan altas temperaturas, la presión de radiación puede superar las fuerzas gravitacionales provenientes del planeta. Si esto ocurre, una columna de gas es levantada en dirección vertical por sobre el planeta, la cual bloqueará una fracción de la radiación estelar a longitudes de onda de $\sim \mu\text{m}$, proyectando una sombra visible en esa frecuencia. Esta sombra se mueve a la velocidad orbital del planeta.

La detección de sombras de este tipo en observaciones de discos protoplanetarios podrían servir como indicadores indirectos de la presencia de planetas en formación. El ejemplo más emblemático es la sombra detectada en HD161942, la cual sugiere que se encuentra un planeta orbitando cerca de la estrella, específicamente a ~ 13 au. Nuestro modelo sugiere programar futuras observaciones de luz de scattering a $\sim 1 \mu\text{m}$ de HD161942 a fin de detectar este potencial planeta en formación.

Otro sistema interesante para desarrollar futuras observaciones es el planeta detectado en PDS70, el cual por sus características debería levantar una columna de gas en sus inmediaciones y potencialmente proyectar una sombra observable en luz scattering. Para una referencia completa sobre lo expuesto aquí, consultar Montesinos et al. (2021).

REFERENCIAS BIBLIOGRÁFICA

- ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, *ApJL*, 808, L3, doi: 10.1088/2041-8205/808/1/L3
- Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, *ApJ*, 869, L41, doi: 10.3847/2041-8213/aaf741
- Avenhaus, H., Quanz, S. P., Schmid, H. M., et al. 2014, *ApJ*, 781, 87, doi: 10.1088/0004-637X/781/2/87
- Bertrang, G. H.-M., Avenhaus, H., Casassus, S., et al. 2018, *MNRAS*, 474, 5105, doi: 10.1093/mnras/stx3052
- Bertrang, G. H. M., Flock, M., Keppler, M., et al. 2020, arXiv e-prints, arXiv:2007.11565. <https://arxiv.org/abs/2007.11565>
- Christiaens, V., Cantalloube, F., Casassus, S., et al. 2019, *ApJL*, 877, L33, doi: 10.3847/2041-8213/ab212b
- Debes, J. H., Poteet, C. A., Jang-Condell, H., et al. 2017, *ApJ*, 835, 205, doi: 10.3847/1538-4357/835/2/205
- Dullemond, C. P., Dominik, C., & Natta, A. 2001, *ApJ*, 560, 957, doi: 10.1086/323057
- Dullemond, C. P., Juhasz, A., Pohl, A., et al. 2012, RADMC-3D: A multi-purpose radiative transfer tool, *Astrophysics Source Code Library*. <http://ascl.net/1202.015>
- Huang, J., Andrews, S. M., Dullemond, C. P., et al. 2018a, *ApJL*, 869, L42, doi: 10.3847/2041-8213/aaf740
- Huang, J., Andrews, S. M., Pérez, L. M., et al. 2018b, *ApJL*, 869, L43, doi: 10.3847/2041-8213/aaf7a0
- Hubeny, I. 1990, *ApJ*, 351, 632, doi: 10.1086/168501
- Keppler, M., Benisty, M., Müller, A., et al. 2018, *A&A*, 617, A44, doi: 10.1051/0004-6361/201832957
- Long, F., Pinilla, P., Herczeg, G. J., et al. 2018a, *ApJ*, 869, 17, doi: 10.3847/1538-4357/aae8e1
- Long, Z. C., Akiyama, E., Sitko, M., et al. 2018b, *ApJ*, 858, 112, doi: 10.3847/1538-4357/aaba7c

- Marino, S., Perez, S., & Casassus, S. 2015, ApJL, 798, L44, doi: 10.1088/2041-8205/798/2/L44
- Masset, F. S. 2002, A&A, 387, 605, doi: 10.1051/0004-6361:20020240
- Mihalas, D., & Weibel Mihalas, B. 1984, Foundations of radiation hydrodynamics
- Montesinos, M., Cuadra, J., Perez, S., Baruteau, C., & Casassus, S. 2015, ApJ, 806, 253, doi: 10.1088/0004-637X/806/2/253
- Montesinos, M., & Cuello, N. 2018, MNRAS, 475, L35, doi: 10.1093/mnras/sly001
- Montesinos, M., & de Freitas Pacheco, J. A. 2011, A&A, 526, A146, doi: 10.1051/0004-6361/201015026
- Montesinos, M., Perez, S., Casassus, S., et al. 2016, ApJL, 823, L8, doi: 10.3847/2041-8205/823/1/L8
- Montesinos, M., Cuello, Nicolás ; Olofsson, Johan search by orcid ; Cuadra, Jorge ; Bayo, Amelia search by orcid ; Bertrang, Gesa H. -M. search by orcid ; Perrot, ClémentThe Astrophysical Journal, Volume 910, Issue 1, id.31,pp
- Mordasini, C., Alibert, Y., Klahr, H., & Henning, T. 2012, A&A, 547, A111, doi: 10.1051/0004-6361/201118457
- Müller, A., Keppler, M., Henning, T., et al. 2018, A&A, 617, L2, doi: 10.1051/0004-6361/201833584
- Müller, T. W. A., Kley, W., & Meru, F. 2012, A&A, 541, A123, doi: 10.1051/0004-6361/201118737
- Natta, A., Prusti, T., Neri, R., et al. 2001, A&A, 371, 186, doi: 10.1051/0004-6361:20010334
- Olofsson, J., Benisty, M., Le Bouquin, J.-B., et al. 2013, A&A, 552, A4, doi: 10.1051/0004-6361/201220675
- Quanz, S. P., Avenhaus, H., Buenzli, E., et al. 2013, ApJL, 766, L2, doi: 10.1088/2041-8205/766/1/L2
- Stolker, T., Dominik, C., Avenhaus, H., et al. 2016, A&A, 595, A113, doi: 10.1051/0004-6361/201528039

TERCERA PARTE

MANEJO DE DATOS Y ACTUALIZACIÓN ESTADÍSTICA

1

Encontrables, accesibles, interoperables y reutilizables: Algunas consideraciones para la gestión de nuestros datos de investigación.

Danilo Reyes-Lillo

*Centro de Recursos para la Enseñanza y el Aprendizaje (CREA)
dreyes@uvm.cl*

RESUMEN

El desarrollo tecnológico ha permitido que las comunidades académicas tengan un mayor acceso a información científica. Lo anterior no solo implica un mayor acceso a revistas, libros o documentación científico-técnica tradicional, sino que también hoy en día se conciben los datos de investigación como un sub-producto de la actividad científica. Estos datos contribuyen a optimizar los recursos favoreciendo su reutilización y mejorando la transparencia para validar los resultados obtenidos. Por lo tanto, las comunidades académicas deben aprender a gestionar correctamente sus datos para fortalecer el ecosistema de Ciencia, Tecnología e Innovación (CTI), dado que incluso las instituciones públicas están impulsando iniciativas como la Política de Acceso Abierto a la Información Científica y Datos de Investigación de la Agencia Nacional de Investigación y Desarrollo (ANID), en donde una correcta publicación de los datos será una exigencia para proyectos de investigación con financiamiento público.

Palabras clave: Datos de investigación, Acceso Abierto, Políticas de Investigación.

Introducción

El desarrollo de las Tecnologías de la Información y la Comunicación (TICs) y la transición de trabajar en computadores aislados a computadores en red, junto con distribuir archivos de forma remota, ha generado un cambio revolucionario en muchos ámbitos de la actividad humana y los procesos de investigación científica no se han visto ajenos a estas

circunstancias. La transformación del papel (lo impreso) hacia lo digital ha provocado un importante cambio en el ámbito del acceso a la información científica pasando esta a denominarse la “revolución del acceso” (Suber, 2015). Por otra parte, no solo el desarrollo tecnológico, sino también la concepción política del acceso a la información académica ha impulsado movimientos como el Open Access (OA) o Acceso Abierto, donde se plantea la necesidad de eliminar las barreras que impiden un acceso fluido a estos contenidos a través de Internet: la barrera económica, por el costo de acceso que imponen algunos editores y la barrera jurídica, por las limitaciones que impone el derecho de autor o el copyright (Abadal, 2013).

Si bien al hablar de acceso abierto a la documentación científica, comúnmente se relaciona con el acceso a artículos de investigación, hoy en día la preocupación y el interés por la preservación y el intercambio de los datos de investigación es cada vez más creciente por parte de las comunidades académicas (González et al, 2013). El desarrollo tecnológico permite el intercambio de artículos, pero también existe la posibilidad de transmitir los datos crudos (raw data) resultantes de un proceso de investigación académica. Si bien este proceso se ha dado hace mucho tiempo de manera no formalizada (a través de correos electrónicos, por ejemplo), las instituciones vinculadas a la producción de información científica están buscando los mecanismos para formalizar y estructurar los protocolos de preservación y difusión de datos de investigación, dado que puede fortalecer e incentivar su reutilización para obtener nuevos y mejores resultados de investigación.

¿Qué son los datos de investigación?

Establecer una definición concreta y precisa sobre los datos de investigación es un desafío no menor. Debido a que el propio proceso investigativo trasciende todas las disciplinas, la producción de datos también se vincula con todas las áreas del conocimiento y también con los diferentes actores que participan del proceso (científicos, universidades, centros de investigación, instituciones financiadoras, etc.).

En términos muy generales, Couto Corrêa (2016) los define como “la información registrada o producida mediante cualquier forma o medios durante el transcurso de una investigación” (p. 27).

Asimismo, a pesar de que, en una primera lectura, el concepto dato puede relacionarse con el ámbito cuantitativo, los datos de investigación pueden ser muy diversos y resultantes de cualquier proceso investigativo. Peset y González (2017) señalan que los datos son declaraciones, cualitativas o cuantitativas, o números que no debieran ser factuales, o al menos se asume que no lo son, y pueden ser crudos/primarios o derivados de los datos primarios, pero que cuentan con la característica de que aún no han sido interpretados o analizados mediante algún proceso metodológico que permita establecer conclusiones con base en los mismos.

De esta manera, es posible identificar datos de tipo numérico, textual, descriptivo o visual, así como pueden estar disponibles en formato impreso o en formato digital. Por lo tanto, los datos de investigación pueden ser datos crudos estructurados o no estructurados, registros estadísticos, gráficos, muestras biológicas o ambientales, fotografías, videos, notas de campo, transcripción de entrevistas, entre muchos otros posibles documentos que se obtienen de un proceso de investigación. En síntesis, son todas aquellas evidencias que deja un investigador para entregar validez a las conclusiones resultantes de su proceso investigativo.

Si bien se hace difícil acotar con precisión el concepto, dado que existen ciertos elementos que aún son discutidos por la comunidad académica sobre una definición certera, Torres-Salinas et. al. (2012) señalan que:

la que más consenso aúna, al haber sido adoptada por entidades como los National Institutes of Health (NIH) de Estados Unidos o la OECD (2007), es la que considera datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar resultados de la investigación que se realiza. Además de estas características también se especifica que deben provenir de una fuente única y deben ser difíciles o imposible de obtener de nuevo por ser propios de un momento o circunstancias irrepetibles de una forma exactamente igual (p. 175).

En ese sentido, aunque hasta hace poco tiempo este tipo de datos no era considerado por la comunidad académica, dado que se limitaban a compartir artículos o libros de revistas, la relevancia de los datos de investigación ha resultado creciente porque son “la materia prima generada a partir de la toma/recogida/observación” de los fenómenos y sucesos de la realidad... considerados como un subproducto que podía servir como punto de partida para avanzar en nuevos estudios” (Couto Corrêa, 2016, p.28).

Beneficios de los datos de investigación

Esta reconsideración de los datos de investigación como un subproducto del cual disponen las comunidades académicas para su aplicación, nos hace pensar en que los investigadores podrían ser los principales beneficiarios, dado que podrían reutilizar los datos para sus propias investigaciones. Sin embargo, no son los únicos, dado que las empresas, organismos de salud o instituciones públicas se pueden beneficiar de igual manera al obtener acceso a información obtenida de procesos de investigación y que se preservan a través de un proceso de gestión.

Por otra parte, la CEPAL (2016) señala que los beneficios de realizar una correcta gestión de los datos de investigación son principalmente los siguientes:

- Uso eficiente de recursos públicos, ya que al financiar investigaciones y disponer los datos, los gobiernos permiten el crecimiento económico y se traduce en un beneficio social.
- Mayor transparencia, ya que aumenta la confianza pública y la posibilidad de validar los resultados de investigación.
- Incrementa la posibilidad de asegurar que los datos que están siendo preservados y sean correctamente accesibles, inteligibles, evaluables y usables por otros.
- Aumenta el estándar de la investigación, ya que resguarda buenas prácticas en el ámbito de la ciencia.
- Los resultados de investigación pueden ser evaluados, validados, replicados o reproducidos, aumentando en su validez y confiabilidad.
- Se contribuye en el ámbito de la colaboración científica y la generación de redes entre investigadores.

De esta manera, el compartir datos de investigación tiene múltiples beneficios, dado que permite obtener un mayor rendimiento a los fondos públicos invertidos en la ciencia, generando un mayor uso y aprovechamiento de los recursos. Por lo tanto, la disposición de los datos contribuye a posibles reutilizaciones de los datos para nuevos estudios, ahorro de costos para evitar la financiación de proyectos duplicados en términos de toma de datos y mayor transparencia para contribuir en la evaluación de los resultados obtenidos por los investigadores (Torres-Salinas et. al., 2012).

Situación de datos de investigación en Chile

Entre las comunidades científicas, académicas y también entre equipos de bibliotecarios en Chile, el interés por los datos de investigación ha sido creciente, sobre todo en el marco del intento por garantizar un mayor y más equitativo acceso a la información científica.

Tal es el fenómeno, que la propia Agencia Nacional de Investigación y Desarrollo (ANID) ha planteado la relevancia que tiene garantizar el acceso a información científica financiada con fondos públicos para promover e incentivar la investigación científica, un problema que preocupa a muchos organismos internacionales que financian proyectos de Ciencia, Tecnología e Innovación (CTI).

Por una parte, el objetivo de libre difusión de los resultados de investigación que son publicados en revistas académicas y que son financiados con fondos públicos, se encuentra en conflicto con los modelos de comercialización de las publicaciones, donde las editoriales transan el conocimiento resultante como cualquier bien o servicio. Por lo tanto, las propias instituciones que producen los resultados de investigación deben pagar suscripciones a un alto costo para acceder a las publicaciones académicas, lo que se contrapone con los objetivos del financiamiento de proyectos de investigación.

Por otro lado, al existir estas barreras económicas de acceso a la información científica, inevitablemente se generan asimetrías entre investigadores, instituciones e incluso entre países. En el plano nacional, uno de los objetivos es reducir las brechas que existen en materia de acceso abierto entre los diversos actores que contribuyen al ecosistema científico.

Por lo tanto, en el año 2020 la ANID ha impulsado una “Propuesta de Política de acceso abierto a la información científica y a datos de investigación financiados con fondos públicos”, incluso sometiéndola a consulta ciudadana a mediados de año, donde se aborda no solo el acceso a publicaciones sino también a los datos de investigación, los cuales define como:

todos aquellos que se generan a lo largo del desarrollo del proyecto de investigación, con o sin tratamiento de por medio, y que sirven de soporte a los resultados del proyecto de investigación. Se incluyen en este conjunto los datos personales, que deberán ser tratados siguiendo las disposiciones reglamentarias y legales vigentes para su debida protección (p. 7).

Asimismo, de acuerdo a los datos de investigación, la propuesta señala que la política se basará en los principios FAIR (en inglés es la sigla de Findability, Accessibility, Interoperability, and Reusability), lo que supone que los datos de investigación deben ser encontrables, accesibles, interoperables y reusables.

De esta forma, el documento de la ANID se refiere específicamente a la obligatoriedad de publicar los datos obtenidos que tendrán los investigadores y las instituciones que se adjudiquen fondos públicos:

Todos los beneficiarios y las beneficiarias de instrumentos con convenios vigentes con la ANID deberán presentar un plan de gestión de datos de investigación, acorde con las especificidades propias del área disciplinar y, en todos los casos, un plan para garantizar la disponibilidad pública de dichos datos (p. 9).

Por otra parte, también se señala que las personas que se adjudiquen fondos deberán depositar los datos en el repositorio de la ANID o en un repositorio propio de la institución

a la que se adscribe la persona responsable del proyecto. Asimismo, estos repositorios deben cumplir todos los estándares técnicos que garanticen el acceso y el resguardo de los documentos o datasets.

En consecuencia a la emergente política de acceso a la información científica, las comunidades académicas e investigadoras a nivel nacional tendrán que desarrollar las habilidades necesarias para poder diseñar un plan de gestión de sus propios datos de investigación, junto con tener el apoyo necesario para disponer en repositorios digitales la información recolectada en su proceso de investigación.

Rol de los investigadores en la gestión de datos de investigación

A raíz de lo expuesto previamente, los investigadores tendrán que adaptarse al nuevo escenario y, adicionalmente a la producción de sus resultados de investigación, tendrán que desarrollar un plan de gestión de datos junto con disponerlos en repositorios digitales, con base en los principios FAIR (Findable, accesible, interoperable and reusable). Estos principios fomentan la aplicación de buenas prácticas en la gestión de datos de investigación centrándose en cuatro principios que componen sus siglas: encontrables, accesibles, interoperables y reusables.

- **Encontrables:** se refiere a utilizar todas las tecnologías posibles (identificadores y metadatos descriptivos) que permitan encontrar fácilmente los datos de investigación.
- **Accesibles:** se refiere a que las tecnologías y protocolos funcionen con la mayor apertura y universalidad posible, garantizando accesibilidad.
- **Interoperables:** se refiere a que sean descritos con estándares que permitan la comunicación con otros sistemas.
- **Reutilizables:** se refiere a que los datos cuenten con los atributos necesarios para ser reutilizados y muestren con claridad las licencias de uso.

Por lo tanto, con base en estos principios, los investigadores deben realizar una adecuada gestión de datos considerando el ciclo de estos mismos, donde se exponen como etapas principales el establecer el plan de gestión de datos; la asignación de metadatos; los estándares, almacenamiento y copias de seguridad; el acceso y utilización, licencias y uso ético (Figura 1).



Figura 1: Ciclo de los datos científicos
Fuente: CRUE - Red de Bibliotecas REBIUN (2016)

En todos estos procesos, los investigadores deben ver como alianza estratégica sus respectivas bibliotecas, ya que pueden recibir la asesoría profesional adecuada en todas la etapas que componen el ciclo de los datos de investigación.

Asimismo, para que un investigador pueda elaborar un buen plan de gestión de datos, la REBIUN (2016) sugiere considerar los 10 siguientes puntos para obtener resultados óptimos:

1. Considerar los requerimientos de la institución que financia la investigación.
2. Identificar los datos según tipología, volumen, formatos y archivos implicados.
3. Definir al organización de los datos, considerando el control de versiones y formatos de presentación.
4. Explicar cómo se documentan los datos.
5. Describir los procesos que aseguran la calidad de los datos.

6. Generar una estrategia de almacenamiento y de preservación, donde toma relevancia el repositorio,
7. Definir las políticas de datos del proyectos, donde se abordan elementos de propiedad intelectual o tratamiento de datos sensibles.
8. Definir una estrategia de difusión de los datos de investigación.
9. Asignar roles y responsabilidades a todas las personas que participan del proyecto.
10. Establecer un presupuesto realista, considerando personal, software y hardware.

De esta manera, podemos establecer las primeras directrices para que los investigadores pueden tener un rol activo en la gestión de sus datos de investigación en sus propios proyectos.

Si bien aún en Chile este es un tema incipiente, las universidades, centros de investigación y la Agencia Nacional de Investigación y Desarrollo están tomando las primeras iniciativas para fortalecer el acceso a la información científica, también considerando los datos de investigación como un importante subproducto que puede fortalecer y beneficiar el ecosistema de Ciencia, Tecnología e Innovación (CTI).

REFERENCIAS BIBLIOGRÁFICAS

- Abadal, E. (2013). Acceso abierto a la ciencia. Editorial UOC. <https://elibro.net/es/lc/bibliouvm/titulos/56586>
- ANID. (2020). Propuesta de Política de acceso abierto a la información científica y a datos de investigación financiados con fondos públicos de la ANID. https://s3.amazonaws.com/documentos.anid.cl/estudios/Politica_acceso_a_informacion_cientifica_version_final_26-05-2020.pdf
- CEPAL. (2016). Gestión de Datos de Investigación: Algunos conceptos básicos. <https://www.cepal.org/es/notas/gestion-datos-investigacion-algunos-conceptos-basicos>
- Couto Corrêa, F. (2016). Gestión de datos de investigación. Editorial UOC. <https://elibro.net/es/lc/bibliouvm/titulos/114200>
- CRUE – Red de Bibliotecas REBIUN. (2016). El ciclo de los datos científicos. <https://rebiun.xercode.es/xmlui/handle/20.500.11967/69>
- González, L., Saorín, T, Ferrer-Sapena, A., Aleixandre-Benavent, R. & Peset, F. (2013). Gestión de datos de investigación: infraestructuras para su difusión. *El Profesional de la Información*, 22(5). pp. 415-423. <http://dx.doi.org/10.3145/epi.2013.sep.06>
- Peset, F. y Millán González, L. (2017). Ciencia abierta y gestión de datos de investigación (RDM). Ediciones Trea. <https://elibro.net/es/lc/bibliouvm/titulos/117484>
- Suber, P. (2015). Acceso abierto. CLACSO. <https://elibro.net/es/lc/bibliouvm/titulos/39283>
- Torres-Salinas, D., Robinson García, N., & Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación: Introducción al data sharing. *El Profesional de la Información*, 2012, 21(2). pp. 173-184. <http://dx.doi.org/10.3145/epi.2012.mar.08>

2

Estadística Descriptiva: El primer paso de toda investigación

Catalina León Martínez

*Escuela de Ciencias
cleon@uvm.cl*

RESUMEN

Cuando nos vemos enfrentados al trabajo de investigación, debemos abordar una serie de pasos antes de encontrarnos cara a cara con los datos. Sin embargo, una vez resueltos aquellos, debemos preguntarnos ¿cómo organizarlos y analizarlos de modo de obtener la mayor información posible de ellos? En muchas ocasiones la premura por dar respuesta a la pregunta que originó nuestra investigación nos hace acercarnos rápidamente a los métodos asociados a la estadística inferencial y olvidar el primer paso: describir la información con que contamos. Un acercamiento exploratorio adecuado de los datos nos permitirá tener mayor claridad acerca de los métodos más eficientes para cumplir con nuestro objetivo inicial o incluso, podría sugerirnos la modificación de éste. El objetivo del presente artículo de divulgación es dar una mirada general a los distintos aspectos que debemos considerar al momento de describir las variables en estudio, así como a las herramientas con las que contamos para ello. En ningún caso pretende ahondar en cálculos, fórmulas o procedimientos, ya que en la actualidad las múltiples tecnologías existentes hacen innecesario este tratamiento. Más bien, tratará de abordar estos resultados de manera crítica, resaltando la necesidad de ver toda la información en conjunto para obtener la mejor descripción posible.

Palabras clave: Variables, distribución de frecuencias, medidas de tendencia central. medidas de dispersión.

Introducción

El presente artículo de divulgación nace bajo el alero de la 2° Jornada de investigación y actualización interna UVM, a realizarse en enero del 2022. El texto pretende dar una visión global de los diferentes aspectos que se deben considerar para un análisis estadístico descriptivo previo a la utilización de metodologías más específicas. Se trata, por lo tanto, de un artículo orientado a quien no tiene mayor formación estadística, pero comienza a introducirse en el mundo de la investigación. El artículo está dividido en 4 partes: la primera aborda el tema de las variables y su clasificación, la segunda, resume los aspectos a describir en un conjunto de datos y los estadísticos pertinentes para ello. En la tercera parte, se hace una pequeña mención del análisis bivariado (dos variables) y la cuarta y última parte, contiene las conclusiones de este trabajo. La idea que da origen a este artículo es contribuir a la formación de una cultura estadística que nos permita relacionarnos adecuadamente con la inmensa información que contamos en la actualidad. Se ha intentado dar varios ejemplos, la mayoría de ellos tomados de registros personales de calificaciones en los cursos donde he impartido docencia, a modo de acercar los conceptos al lector.

Parte 1: Variables

Una vez determinada la pregunta de interés que da origen a una investigación, se requiere identificar las variables que pueden contribuir a darle respuesta. Entenderemos por variables a las características en estudio de los elementos de una población (unidades de observación) y llamaremos datos a los valores observados de tales variables. Es importante mencionar que existen distintos tipos de variables y que resulta de gran importancia identificar aquellas con las que estamos trabajando.

Consideremos un grupo de personas de las cuales hemos registrado sus edades y ciudad de nacimiento. Sin entrar en detalles, es claro que podríamos calcular el promedio de las edades de tales personas, pero ¿podemos calcular el promedio de las ciudades de nacimiento? Evidentemente no. Esta situación nos permite constatar, por una parte, que no todas las variables responden a una misma naturaleza (no son del mismo tipo) y por otra, que las herramientas estadísticas utilizadas para describir tales variables dependen justamente del tipo de variable con el que estamos trabajando.

Si bien existen diferentes criterios para clasificar una variable, hemos optado por una clasificación sencilla, pero que da cuenta de importantes distinciones. El siguiente esquema la resume.

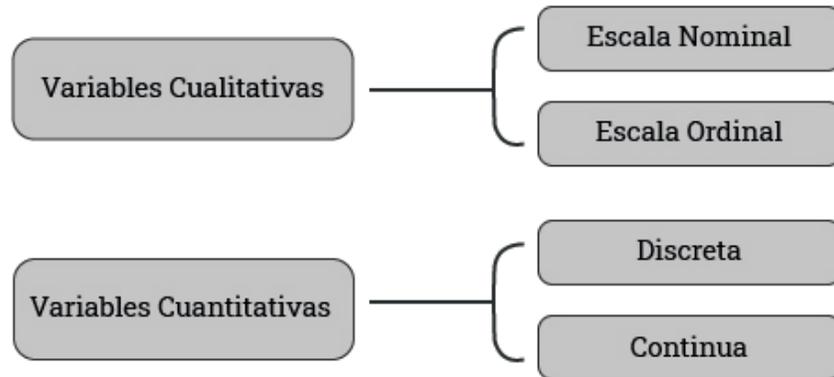


Figura 1: Clasificación de variables

Repasemos cada una de las categorías:

- a) Variables Cualitativas: son aquellas cuyos valores son un conjunto de cualidades no numéricas a las que se les suele llamar categorías, modalidades o niveles. Por ejemplo: sexo (mujer, hombre), estado civil (soltero, casado, divorciado, viudo).
- b) Variables Cuantitativas: son aquellas cuyos posibles valores son numéricos. Por ejemplo, estatura, peso, edad, temperatura, etc.

Las variables cualitativas a su vez pueden clasificarse según la escala de medición en nominales u ordinales:

- a) Nivel de medición nominal: los datos consisten exclusivamente en nombres, rótulos o categorías y se clasifican en subconjuntos mutuamente excluyentes. Ejemplo: sexo (hombre, mujer), color de ojos, etc.
- b) Nivel de medición ordinal: en este caso, a diferencia del anterior, se pueden establecer relaciones de orden entre las categorías según el grado en que posean la característica. Ejemplo: grupos socioeconómicos, desempeño en un curso (insuficiente, suficiente, bueno, muy bueno, excelente), talla (S, M, L, XL) etc.

Por otra parte, las variables cuantitativas pueden clasificarse según sus posibles valores como:

- a) Discretas: son aquellas que pueden tomar un conjunto finito o infinito contable de valores. Por ejemplo, número de hijos, número de personas en una fila, etc.
- b) Continuas: son aquellas que pueden tomar un conjunto infinito no contable de valores. Por ejemplo: duración de una batería, edad, temperatura, etc.

Parte 2: Describiendo el comportamiento de la variable

¿Por qué es importante organizar y describir los datos? El ordenamiento y el análisis adecuado de los datos u observaciones registradas pueden llevarnos a detectar aspectos de la realidad que de otra forma permanecerían ocultos. No se trata, por tanto, de un análisis opcional, sino, muy por el contrario, todo estudio debe comenzar realizando un análisis exploratorio de los datos, desde donde emergerá información que iluminará los métodos más adecuados en un análisis posterior... Pero ¿qué aspecto describiremos?

1. Distribución de frecuencias.

En primer lugar, exploraremos cómo están distribuidos los datos: ¿Entre qué valores se concentra la mayor cantidad de datos?, ¿Dónde hay menor concentración?, ¿Se trata de una distribución simétrica?, ¿Se asemeja a alguna distribución teórica?, etc.

Para describir la distribución de las frecuencias contamos básicamente con dos herramientas: tablas y gráficos.

a) Tablas de frecuencia

Una tabla de frecuencia tiene la siguiente estructura:

Título:

Nombre de la Variable	Tipo de Frecuencia
Categorías de la Variable	Frecuencias Observadas

Figura 2: Estructura de una tabla de frecuencia

En la literatura suelen identificarse cuatro tipos de frecuencias: frecuencia absoluta (que corresponde simplemente a la cantidad de datos en cada una de las categorías), frecuencia relativa (expresada por lo general en porcentaje) y las respectivas frecuencias acumuladas. También suele hacerse la diferenciación en tablas de frecuencias simples (cuando en la primera columna consideramos puntualmente todos los valores posibles de la variable) y la tabla de datos agrupados (cuando debido a la naturaleza continua de la variable se agrupan sus valores en rangos o intervalos). Para más detalle sobre la construcción de cada una de ellas puede consultar Estadística Descriptiva (Montero, 2007).

Notemos que si la variable en estudio es una variable cualitativa nominal, las frecuencias acumuladas no tienen ningún sentido, pues no existe una relación de orden entre las categorías.

Si bien las tablas de frecuencias nos entregan información importante en relación con la distribución de las frecuencias, la visualización de ellas a partir de gráficos nos permite un análisis más global.

Gráficos

¿Qué gráficos nos permiten visualizar frecuencias? Existen varias alternativas, sin embargo, el gráfico más apropiado nuevamente estará determinado por la naturaleza de la variable en estudio. Si la variable es cualitativa nominal sin muchas categorías distintas, probablemente la mejor opción es el gráfico circular, pues permite visualizar las frecuencias como la parte de un todo facilitando una lectura rápida de éste. Si la variable es cuantitativa discreta sin muchas categorías distintas, un gráfico de barras simple podría ser también la opción. Ahora bien, si la variable en estudio es continua, la herramienta gráfica correspondiente es el histograma. El histograma consiste en una serie de rectángulos adyacentes cuyas áreas son proporcionales a la frecuencia del intervalo sobre el cual se ubica. Si los intervalos son de igual amplitud, los rectángulos tendrán una altura proporcional a la frecuencia correspondiente.

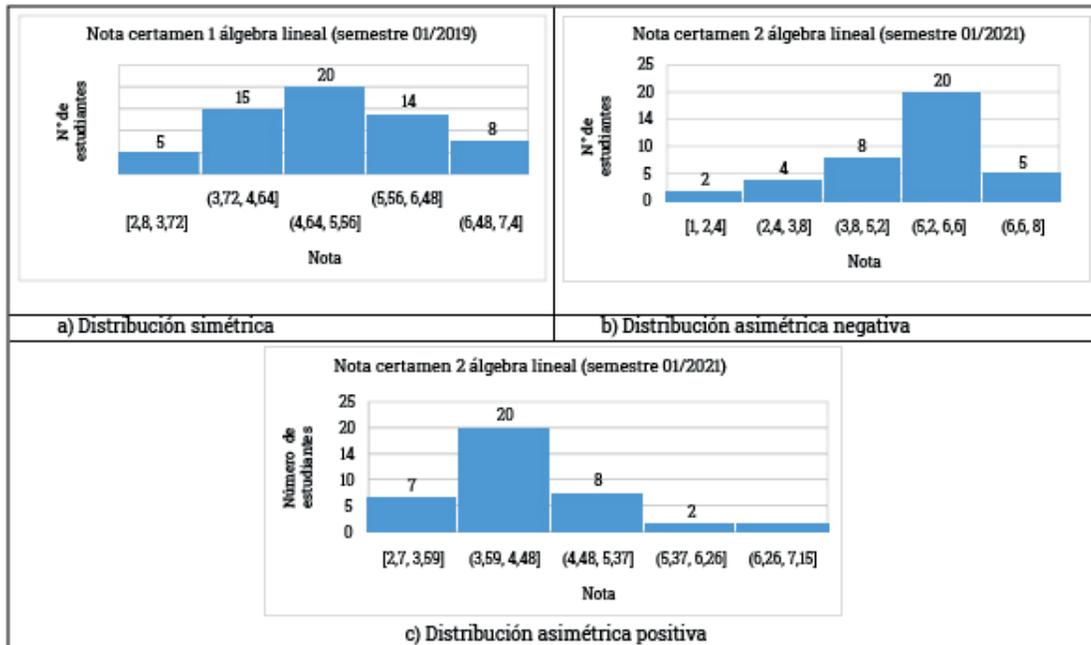


Figura 3: Ejemplos de distribución simétrica, asimétrica negativas y asimétrica positiva

El análisis de un histograma puede llevarnos a concluir acerca de algún comportamiento sistemático de los datos. Por ejemplo, podemos distinguir, entre otras, tres situaciones:

- La distribución es asimétrica negativa
- La distribución es simétrica
- La distribución es asimétrica positiva

A modo de ejemplo, los histogramas de la figura 3 representan el rendimiento en distintas evaluaciones de un curso de álgebra lineal. En particular, en la figura 3 a) podemos observar que existe un comportamiento simétrico en las notas, mientras que en las figuras 3 b) y 3 c) se aprecia un comportamiento asimétrico negativo y asimétrico positivo, respectivamente.

Dada la importancia que reviste para distintos análisis la simetría de la distribución, el histograma nos da una primera aproximación acerca de la plausibilidad de considerar este supuesto.

2. Posiciones relativas (Cuantiles)

Un buen complemento del punto anterior, son los denominados cuantiles de una distribución. En la actualidad muchas decisiones se toman en base a esta medida descriptiva. Por ejemplo, es común escuchar afirmaciones del tipo: “la beca de estudio beneficiará a los tres primeros cuantiles” o “se aumentarán los impuestos al último decil de la población”. En ambos casos la variable considerada es el ingreso económico de una familia aunque no se encuentra declarada explícitamente. En términos generales las medidas de posición se denominan cuantiles y dividen la distribución de los datos en una cierta cantidad de partes iguales. El nombre específico se deriva de la cantidad de partes en la que se divide la distribución:

a) Cuantiles:

Divide la distribución de los datos en cuatro partes iguales una vez ordenados según su magnitud. Existen 3 cuantiles y se denotan con la letra Q.



b) Quintiles:

Divide la distribución de los datos en 5 partes iguales una vez ordenados según su magnitud. Existen 4 quintiles y se denotan con la letra F.



c) Deciles:

Divide la distribución de los datos en 10 partes iguales una vez ordenados según su magnitud. Existen 9 deciles y se denotan con la letra D.



d) Percentiles:

Divide la distribución de los datos en cien partes iguales una vez ordenados según su magnitud. Existen 99 percentiles y se denotan con la letra P. Note que todas las otras medidas de posición pueden expresarse como percentiles. Por ejemplo, Q_1 corresponde a P_{25} , D_8 corresponde a P_{88} , F_3 corresponde a P_{60} , etc.

1%	1%	1%	1%	1%	...	1%	1%
Mín	P_1	P_2	P_3	P_4	P_5		Max

Una pequeña nota: el concepto de cuantil es de vital importancia al momento de analizar si la distribución empírica de los datos se puede aproximar mediante una distribución teórica, pues la idea es justamente comparar los cuantiles empíricos con los teóricos, de modo que, si no existen grandes diferencias, podría asumirse esa distribución.

3. El centro de la distribución (Medidas de Tendencia Central)

En una primera etapa, será probablemente de interés resumir todos los datos en un solo valor. Resulta evidente que un buen representante debe reflejar, de alguna manera, el centro de la distribución. Pero ¿qué consideraremos como centro?

a) La media aritmética (promedio)

Probablemente la idea de centro más extendida es la que se refiere a la media aritmética o simplemente media o promedio. Este estadístico, para la mayoría, tiende a ser la medida de tendencia central por excelencia y si bien tiene características que la hacen una buena medida de resumen, también es cierto que debemos analizar cuidadosamente si cumple aquello que promete: representar los datos obtenidos.

Aunque, como hemos dicho, no entraremos en los detalles de cálculo dada la gran variedad de recursos tecnológicos que se pueden ocupar actualmente de esta tarea, no está de más recordar que el promedio corresponde a la suma de todas las observaciones dividida por el número total de ellas, por tanto, es posible calcularla únicamente en variables cuantitativas.

Si nos referimos la media aritmética de la población la denotaremos por μ , mientras que si se trata de la media aritmética de la muestra utilizaremos \bar{x} . A lo largo de este artículo consideraremos que nuestros datos son muestrales, por lo que haremos referencia a los estadísticos (características numéricas de la muestra) y no a los parámetros correspondientes (características numéricas de la población).

¿Cuál es la idea de centro que hay detrás de esta medida? Si pudiéramos un punto de apoyo en la posición de la media en el histograma, este quedaría equilibrado como lo muestra la figura 4. Dicho de otra manera, si pudiéramos “repartir equitativamente” la variable en cuestión entre todos los participantes de la muestra, a cada uno le correspondería el valor de la media. Por ejemplo, si consideramos que, según el INE, el ingreso laboral promedio mensual en Chile en el año 2020 fue de \$635.134, esto significaría que, si se hiciera un gran “fondo común” con el ingreso de todos los chilenos y se repartiera equitativamente, cada uno recibiría ese monto mensualmente. Visto de esta manera, la media parece no ser muy representativa de nuestra realidad.

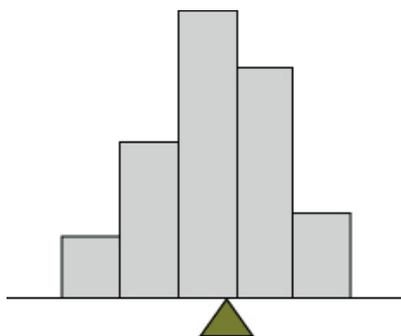


Figura 4: Interpretación de centro según la media aritméticas

Teniendo presente lo anterior, podemos notar que uno de los puntos a favor de la media aritmética es que considera en su cálculo todos los valores observados. Sin embargo, en lo que respecta a su real representatividad debemos tener en cuenta dos aspectos: la presencia de valores extremos (o también denominados outliers) y la dispersión de los datos. Si bien abordaremos este último concepto en el apartado siguiente, podemos referirnos a él de una manera intuitiva en los siguientes ejemplos.

Supongamos que los siguientes conjuntos de datos corresponden a la edad de distintos grupos de personas:

Grupo 1:

5	12	24	35	43	57	62	70	72	80
---	----	----	----	----	----	----	----	----	----

Grupo 2:

39	41	44	44	45	48	50	50	50	50
----	----	----	----	----	----	----	----	----	----

Si obtenemos la media de ambos grupos, podemos corroborar que difieren mínimamente. La edad promedio del grupo 1 es de 46 años, mientras que la del grupo 2 es de 46,1 años. Sin embargo, si tuviéramos que imaginar un grupo de personas cuya edad promedio es de 46 años, probablemente pensaríamos en una situación más bien como la del grupo 2. Cuando la dispersión de los datos es mayor, la media aritmética resulta ser menos representativa.

Ahora bien, consideremos los siguientes casos:

Grupo 3

2	2	2	3	3	3	4	4	5	6
---	---	---	---	---	---	---	---	---	---

Grupo 4

88	2	2	3	3	3	4	4	5	6
----	---	---	---	---	---	---	---	---	---

Las edades de ambos grupos difieren únicamente en un valor. Sin embargo, la edad promedio en el grupo 3 es de 3,4 años, mientras que en el grupo 4 es de 12 años. Claramente este valor no representa en lo absoluto la situación real. ¿Cuál es el problema? En el grupo 2 existe un valor extremo y la media es sensible a este.

Los ejemplos anteriores dan cuenta de la necesidad de considerar conjuntamente las estadísticas descriptivas y no solo como números aislados.

a) La mediana

La mediana corresponde al valor central la distribución una vez que los datos han sido ordenados según su magnitud, de manera que el 50% de los datos son iguales o menores a ella (percentil 50). Notemos que, para el cálculo de la mediana, a diferencia de la media, utilizamos, a lo más, solo dos valores de la distribución (los valores centrales). Sin embargo, es una muy buena opción cuando detectamos valores extremos en la distribución, ya que esta medida resulta mucho más robusta que la media, es decir, no es sensible a la presencia de valores extremos.

Volvamos a uno de los casos anteriores:

Grupo 3

2	2	2	3	3	3	4	4	5	6
---	---	---	---	---	---	---	---	---	---

Grupo 4

88	2	2	3	3	3	4	4	5	6
----	---	---	---	---	---	---	---	---	---

2	2	3	3	3	4	4	5	6	88
---	---	---	---	---	---	---	---	---	----

En el grupo 3 la mediana es 3 años, que corresponde al promedio de los dos valores centrales. En el grupo 4, luego de ordenar los datos según su magnitud, podemos determinar la mediana calculando el promedio de los dos valores centrales de la distribución (3 y 4). La mediana es 3,5 años, es decir, a pesar de existir un valor extremo, la mediana no se ve afectada mayormente.

Recordemos que en el apartado anterior hicimos referencia ingreso promedio mensual de los chilenos. Resulta interesante conocer el ingreso mediano del mismo año y según la misma fuente: \$420.000, cifra que dista bastante de los \$635.134 que se indicaba como promedio...

c) La moda

Finalmente, la última medida de tendencia central que abordaremos es la moda. La moda corresponde al valor de la distribución con mayor frecuencia y NO corresponde necesariamente a la mayoría. La moda puede no existir o no ser única. En este último caso podemos tener una distribución bimodal, por ejemplo. Es importante notar que la moda es la única medida de tendencia central que podemos calcular para variables cualitativas.

4. La dispersión o variabilidad

Como hemos dicho anteriormente, la dispersión o variabilidad es una característica fundamental a la hora de describir un conjunto de datos. Más aún, si no hubiera variabilidad en los datos no tendría sentido realizar un análisis estadístico, pues nuestro objeto de estudio son justamente esas diferencias.

Así como existían distintas ideas de centro, también podemos identificar distintas ideas para determinar la variabilidad o dispersión.

a) La dispersión como rango

Una mirada muy general permite medir la variabilidad de un conjunto de datos como el rango de las observaciones, es decir, la diferencia entre el dato mayor y el dato menor. Claramente esta medida se vería fuertemente afectada por la presencia de un valor extremo, razón por la cual en ocasiones se opta por el rango intercuartílico, es decir, el rango entre el cuartil 3 y el cuartil 1.

b) La dispersión como desviación de la media.

Aunque a lo largo del presente artículo no hemos estudiado en profundidad las fórmulas para el cálculo de los estadísticos correspondientes, en este caso haremos una excepción, no con el fin de introducir alguna dificultad, sino con el objeto de comprender qué estamos midiendo específicamente.

Consideremos medir la variabilidad de los datos como desviación de la media. Parece lógico entonces considerar en primer lugar cuánto se aleja cada una de las observaciones de la media mediante la diferencia:

$$x_i - \bar{x}$$

Ahora bien, sería de utilidad poder resumir esas diferencias mediante un único valor. Sin embargo, la suma de ellas, que resultaría lo más inmediato, no es de utilidad, ya que es posible verificar, según las propiedades de la media que

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Esta igualdad se debe básicamente a que algunos de los resultados de las diferencias son positivos y otros negativos los que finalmente, al sumarse, terminan cancelándose. ¿Cómo podemos evitar esta situación? Si elevamos al cuadrado las diferencias, obtendremos solamente valores positivos y el problema desaparece.

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

Finalmente obtenemos el “promedio” de estas diferencias al cuadrado. ¿Por qué “promedio”? Porque en este caso, por motivos que no profundizaremos en esta oportunidad, se divide por $n-1$ y no por n como sería lo habitual.

¡Hemos deducido la fórmula de la varianza! Esta se denota como s^2

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Pero ¿cuál es la unidad de la varianza? Al elevar al cuadrado las diferencias entre las observaciones y el promedio, la unidad de medida también se eleva al cuadrado, es decir la unidad de la varianza corresponde al cuadrado de la unidad de medición original de los datos. Esto dificulta su interpretación, porque ¿qué quiere decir que la varianza de las edades sea 42 años²? ¿O que la varianza de los pesos de las personas sea de 63 kg²?

Para volver a la unidad original podemos calcular la raíz cuadrada del resultado anterior. Esta nueva medida es lo que se conoce como desviación estándar o típica y se denota con la letra s .

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Es importante resaltar que tanto la varianza como la desviación estándar consideran la variabilidad como desviaciones de la media. ¿Serán adecuadas entonces cuando la media no es representativa?

c) Medidas des dispersión relativa

Como hemos visto en los casos anteriores, la unidad de medida puede hacer la diferencia al momento de interpretar los valores obtenidos: sin embargo, ¿podemos comparar la dispersión en una variable como estatura de una persona (medida en metros) con la variable peso (en kg)?

Supongamos que obtenemos en el primer caso una desviación estándar de 0,15 m y en el segundo, una desviación de 7kg. ¿Quiere decir esto que al ser mayor la desviación estándar del peso su variabilidad es mayor que la existente en la estatura? Claramente no, pues ¿cómo podríamos comparar metros con kilogramos? Necesitamos, entonces, una medida que no dependa de las unidades de medición y que, por lo tanto, nos permita realizar estas comparaciones.

La medida más común que cumple con esta condición es el Coeficiente de Variación de Pearson. Este relaciona mediante un cociente el valor de la desviación estándar con la media aritmética.

Es decir,

$$CV = \frac{s_x}{\bar{x}}$$

Como vemos, el coeficiente de variación representa el número de veces que la desviación estándar contiene a la media, por lo tanto, mientras mayor sea este coeficiente, mayor dispersión existirá.

d) Una mención especial: Box plot

El box plot o gráfico de cajón con bigotes, es una representación que resume visualmente muchos de los aspectos que hemos considerado. Se le conoce también como el esquema de los 5 número porque considera: máximo y mínimo de la distribución, los cuartiles 1 y 3 y la mediana. La mayoría de los paquetes estadísticos identifican también los valores extremos, hecho fundamental para poder determinar, entre otras cosas, la representatividad de algunas medidas descriptivas.

En el siguiente ejemplo, vemos el resumen de las notas obtenidas en el primer certamen del curso Álgebra lineal durante el primer semestre de año 2021.

Min	1
Q1	2,85
Median	3,5
Q3	3,9
Max	5,1
Mean	3,387179
Outliers	6,8

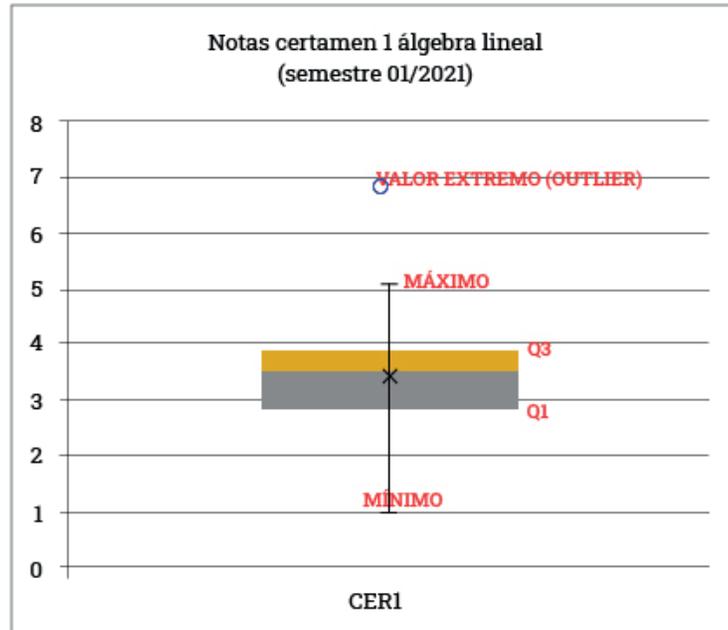


Figura 5: Ejemplo de las componentes de un boxplot

En este caso particular, además de los 5 estadísticos ya nombrado, el gráfico representa con un X la media aritmética de la distribución.

Más allá de los valores numéricos descritos en esta gráfica, el análisis de ella nos permite también tener una idea de la forma de la distribución general de las notas. Notemos que el rango en el que se encuentra el 50% de las notas más bajas es más amplio que el rango en el que se encuentra el 50% más alto. Esto da cuenta de que hay mayor concentración de datos en este segmento y por lo tanto, estaríamos ante una distribución levemente asimétrica negativa.

El gráfico de cajón con bigotes (o box plot) es también una buena herramienta para comparar el comportamiento de una variable en distintos grupos o en distintos momentos. En nuestro ejemplo podríamos estar interesados en saber cómo fue el rendimiento de los estudiantes de Álgebra lineal en los otros dos certámenes del semestre.

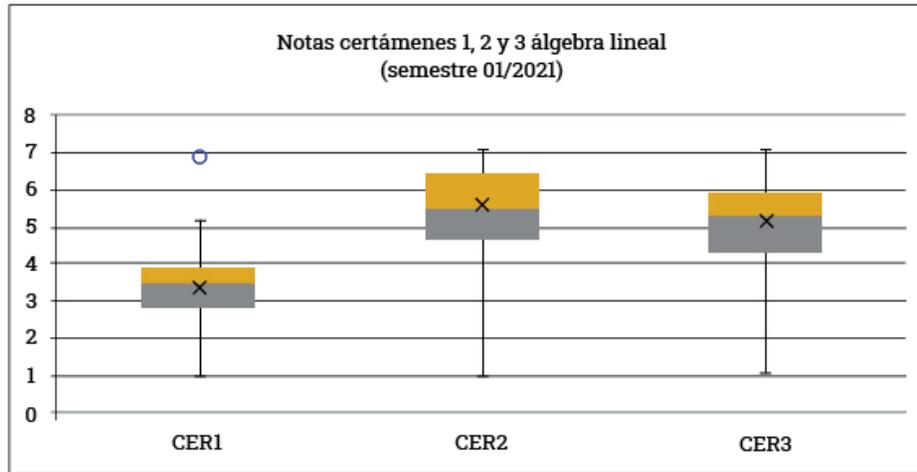


Figura 6: Boxplot para los tres certámenes de álgebra lineal del semestre 01/2021

Según podemos observar en la figura 6, el rendimiento de los estudiantes de álgebra mejoró en las otras evaluaciones. Es posible identificar claramente un aumento tanto de la mediana como de la media, y en general el 50% central de las notas (el cajón), tanto en el certamen 2 como en el certamen 3, está sobre la nota de aprobación (4,0) a diferencia de lo que ocurrió en el certamen 1.

Parte 3: Relacionando dos variables

En muchas ocasiones nuestro interés no se limita al análisis individual de cada una de las variables de nuestro estudio, sino que nos interesa encontrar posibles relaciones entre ellas, y por lo tanto debemos realizarlas conjuntamente.

1. Tablas bidimensionales

Cuando nos vemos enfrentados a un conjunto de datos, la primera prioridad es ordenarlos de alguna manera tal que podamos tener una visión más global de ellos. Así como hablamos de las tablas de frecuencias para describir las frecuencias en el caso univariado, también podemos realizar una tabla bidimensional (o de contingencia) para expresar tanto las frecuencias conjuntas como las marginales de cada variable. Típicamente una tabla de contingencia para una muestra de n datos de las variables X e Y que se clasifican en f categorías de X y c categorías de Y , tiene a siguiente estructura:

X \ Y	B_1	B_2	B_c	$n_{i.}$
A_1	n_{11}	n_{12}	n_{1c}	$n_{1.}$
A_2	n_{21}	n_{22}	n_{2c}	$n_{2.}$
:	:	:		:	:
:	:	:		:	:
A_f	n_{f1}	n_{f2}		n_{fc}	$n_{f.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$		$n_{.c}$	$n_{..}$

Tabla 1: estructura de una tabla de contingencia simétrica negativa y asimétrica positiva

n_{ij} : Frecuencia absoluta conjunta de la categoría i de la variable X y la categoría j de la variable Y

$n_{.j}$: Frecuencia absoluta marginal de la categoría j de la variable Y

$n_{i.}$: Frecuencia absoluta marginal de la categoría i de la variable X

$n_{..}$: n

	Nota Final								
Nota presentación	1,1-1,6	2,1-2,6	3,6-4,1	4,1-4,6	4,6-5,1	5,1-5,6	5,6-6,1	6,6-7,1	Total
1,1-1,6	1								1
2,6-3,1		1							1
3,6-4,1			1	3					4
4,1-4,6			3	4					7
4,6-5,1				3	8	1			12
5,1-5,6						8			8
5,6-6,1							5		5
6,6-7,1								1	1
Total	1	1	4	10	8	9	5	1	39

Tabla 2: Tabla de contingencia para la nota de Presentación y nota de examen de los estudiantes de álgebra lineal

La tabla nos revela información importante: al parecer la nota de presentación y la nota de examen están relacionadas, ya que las observaciones se concentran principalmente en la diagonal. Es decir, a notas de presentación bajas, en general, le corresponden notas de examen bajas y viceversa.

2. Gráfico de dispersión

Si queremos inspeccionar gráficamente la posible relación existente entre dos variables, entonces debemos construir un gráfico de dispersión.

En este tipo de gráfico se representa cada variable en un eje y se observa una nube de puntos donde cada punto corresponde a una observación. El gráfico de dispersión nos permitirá detectar visualmente si existe alguna relación entre las variables y aunque podemos observar un sinfín de situaciones, la figura 7 resume algunos casos que pueden ser de interés.

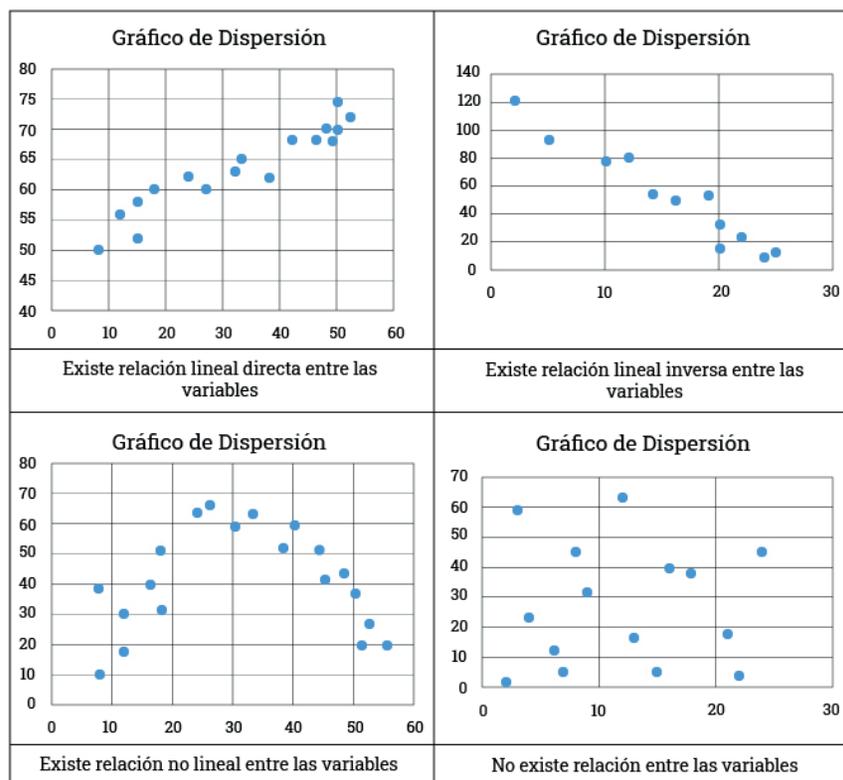


Figura 7: Algunas relaciones detectables en un gráfico de dispersión

Veamos qué ocurre al realizar el gráfico de dispersión entre la nota de presentación y las notas del examen en el curso de Álgebra lineal

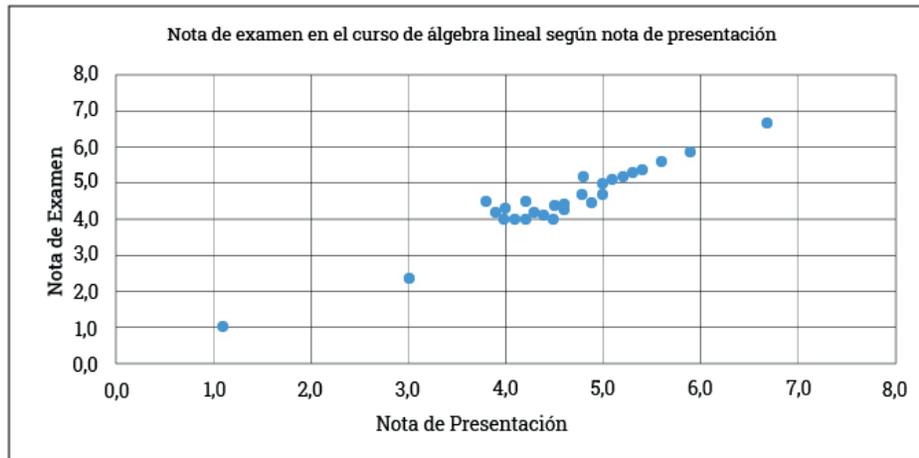


Figura 8: Gráfico de dispersión notas de examen según notas de Presentación

Como vemos en la figura 8, el gráfico de dispersión confirma la relación que habíamos detectado en la tabla de contingencia: existiría, por lo tanto, una relación lineal directa entre la nota de presentación y la nota de examen.

Parte 4: Conclusiones

Probablemente muchos de los conceptos aquí tratados, sino todos, son ya conocidos para el lector. Sin embargo, justamente lo que creemos ya saber no lo miramos, pues no esperamos novedad en ello y perdemos así información importante.

No hay mayor dificultad, especialmente con la tecnología actual, en calcular diversos estadísticos o en construir diversos gráficos. En cualquier programa estadístico, planilla de cálculo o incluso en una calculadora científica, basta introducir datos y nos entregará todos los estadísticos que le solicitemos, ni siquiera es necesario acceder a complejos paquetes estadísticos como hace algunos años atrás. Sin embargo, nada podrá reemplazar la mirada crítica de quien interpreta esa información. No pocas veces me he encontrado con que se confunde la frecuencia de una variable con sus valores y se ha calculado promedio de variables cualitativas, sin provocar ninguna inquietud en quien realiza el estudio. He escuchado también como se pide que el rendimiento de un grupo de trabajadores esté por sobre el promedio, ¡Como si fuera posible que todos los valores estén por sobre la media! He visto también cómo se justifican decisiones a partir del cálculo de promedios, pero no hay ningún análisis de la variabilidad de los datos o de la presencia de valores extremos. ¿Será realmente el promedio un buen representante de la variable en estudio como para basar decisiones en él?

La invitación es a explorar y “estrujar” cada conjunto de datos que analicemos, no solo calculando un listado de medidas, sino analizando cómo ellas se relacionan entre sí y como se iluminan unas a otras. La promesa es que un análisis descriptivo exhaustivo conllevará un análisis posterior mucho más llevadero, ajustado a la realidad y libre de prejuicios previos.

REFERENCIAS BIBLIOGRÁFICAS

- Devore, L. (1998). Probabilidad y Estadística para ingeniería y ciencias (Cuarta Edición). Thomson
- Lind, D., Mason, R., & Marchal, W. (2001). Estadística para Administración y Economía (Tercera edición). McGraw-Hill
- Montero, J. (2007). Estadística Descriptiva. Thomson.
- Sampieri, R., Fernández, C, & Baptista, P. (2004). Metodología de la Investigación (Tercera edición). McGraw-Hill Interamericana.
- Triola, M. (2008). Estadística. Pearson Educación.

3

Usos y abusos en los análisis de regresión y correlación

Adriana Villa-Murillo

*Escuela de Ciencias
adriana.villa@uvm.cl*

RESUMEN

Los análisis de regresión y correlación, en general, son ampliamente empleados como modelos predictivos en los diferentes contextos profesionales. Sumado a esto, el acceso a programas computacionales cada vez más intuitivos desde el punto de vista procedimental, han conllevado al empleo desmesurado de estos tipos de modelos por su aparente sencillez. El objetivo del presente artículo de divulgación es fomentar un pensamiento crítico en el lector en cuanto a si son oportunos (o no) estos modelos según sus objetivos de investigación. Es decir, no se pretende enseñar cómo estimar un modelo de regresión y correlación lineal, sino que se presentan sus principales cimientos con el fin de que el lector decida cuándo es posible y coherente ajustar tales modelos en concordancia con sus objetivos de investigación y naturaleza de los datos.

Palabras clave: Regresión lineal, coeficiente de determinación, correlación lineal..

Introducción

El presente artículo de divulgación aspira dar continuidad a un proceso de alfabetización estadística iniciado en el 2021 mediante los seminarios presentados el texto: La investigación científica desde las metodologías cuantitativa, cualitativa, mixta y sus aspectos éticos³, en el marco de la 1° Jornada de investigación y actualización interna UVM.

³ <http://repositorio.uvm.cl/handle/20.500.12536/1210>

Si el lector aún no conoce tales aportes, le invito cordialmente a visitar el repositorio institucional Space, (Costa S., y otros, 2001) donde podrá actualizar sus conocimientos en cuanto al método estadístico y algunas consideraciones en la metodología de muestreo.

Entonces, en el presente escrito se da un paso más. Se pretende conducir al lector hacia una reflexión en cuanto a los análisis de regresión y correlación como métodos de estimación ampliamente empleados y también sobrevalorados. Para ello, como primer punto (parte 1) se introduce al lector en los principales componentes del modelo de regresión lineal simple, resaltando la importancia del coeficiente de determinación. Seguidamente, en la parte 2, se da paso a los análisis de correlación, todo lo anterior desde un caso de estudio para su mejor comprensión. En la parte 3 resalto, en mi experiencia como estadístico, los usos y abusos que más frecuentemente encontramos en el mundo académico, con el objetivo de establecer las diferencias entre tales metodologías y así propiciar un pensamiento crítico ante futuras aplicaciones. Finalmente, unas breves conclusiones y recomendaciones al lector se manifiestan en la parte 4, junto con unas referencias bibliográficas simples que pueden complementar todo lo aquí expuesto.

Parte 1: Un acercamiento a la regresión lineal.

Los análisis de regresión se encargan de estudiar la relación funcional entre 2 o más variables, es decir, identificar el modelo adecuado con el objetivo de estimar sus parámetros y probar hipótesis con vista a la estimación. Se trata entonces de establecer y probar la relación funcional entre variables con el propósito de hacer predicciones (inferencias) de las variables implicadas bajo una medida de precisión con que esa estimación se hace, pero respetando el dominio de las variables bajo estudio.

Es importante mencionar que el modelo establecido está íntimamente ligado al conocimiento que el investigador posea en el objetivo bajo estudio, dado que de ello dependerá la complejidad y adecuación del modelo. Los modelos de regresión como funciones matemáticas pueden ser de tipo lineal o no, es decir, modelos conformados por el producto de un parámetro y las variables independientes en bloques de sumas, o modelos donde los parámetros pueden estar presentes mediante logaritmos o exponentes de las variables independientes. Entonces, la linealidad de un modelo de regresión hace alusión a la condición “lineal” de los parámetros en la función. El objetivo del presente escrito se enfoca en los modelos de regresión lineal, no solo por su gran aplicabilidad sino porque muchos modelos pueden ser expresados de forma lineal bajo algunas transformaciones matemáticas, lo que comúnmente se llama modelos intrínsecamente lineales.

La manera más sencilla de expresar un modelo de regresión lineal es la que se presenta en la ecuación (1), donde (α, β) son los parámetros a estimar en el modelo, x_i es la variable independiente en el i -ésimo nivel, y_{ij} corresponde a la j -ésima observación en el nivel i -ésimo de la variable dependiente, finalmente, ε_{ij} denota el error aleatorio en los niveles ij correspondientes. Vale la pena resaltar desde ya que en ε_{ij} es donde recaen los supuestos fundamentales que darán o no validez estadística al modelo.

$$y_{ij} = \alpha + \beta x_i + \varepsilon_{ij} \quad (1)$$

Con el objetivo de ir profundizando en el tema, adoptemos un estudio de dominio público⁴, en el cual a 25 individuos se les mide las variables: edad, peso y cantidad de grasa en la sangre. La figura 1 muestra los gráficos de dispersión de tales variables. Note que en ambos casos y, previos conocimientos en el área nutricional, la variable grasas se ha considerado como dependiente. Lo anterior lo podemos expresar matemáticamente, en concordancia con (1), mediante la ecuación (2); pero observe que a priori se está asumiendo una relación lineal que poco se visualiza en la figura 1(b). He aquí la necesidad de la exploración gráfica, antes de plantearnos una relación funcional entre 2 variables.

$$grasas_{ij} = \alpha + \beta(\text{peso})_i + \varepsilon_{ij}$$

$$grasas_{ij} = \alpha + \beta(\text{edad})_i + \varepsilon_{ij} \quad (2)$$

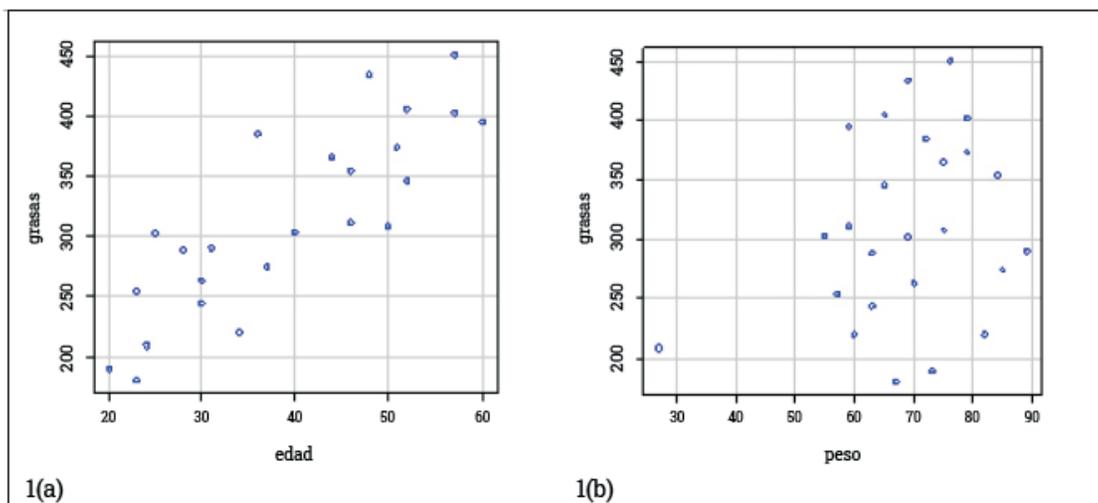


Figura 1: Exploración de las relaciones edad~grasas y peso~grasas

⁴ <http://verso.mat.uam.es/~joser.berrendero/datos/EdadPesoGrasas.txt>

Los análisis de regresión lineal como métodos de estimación lo que coloquialmente hacen es ajustar la nube de puntos a una línea recta mediante valores esperados, en nuestro caso: estimar el valor esperado de grasa para una edad en particular. Retomemos la figura 1(a) pero trazando una línea que intente tocar la mayoría de los puntos (figura 2a). Se debe observar que necesariamente la línea debe pasar por la parte central de la nube de puntos.

¿El centro?: piense por un momento a que se refiere cuando hablamos de la concentración de observaciones en la parte central.

¡Exacto!, al **promedio** y ése es el principio inicial de los análisis de regresión lineal. Me explico a continuación.

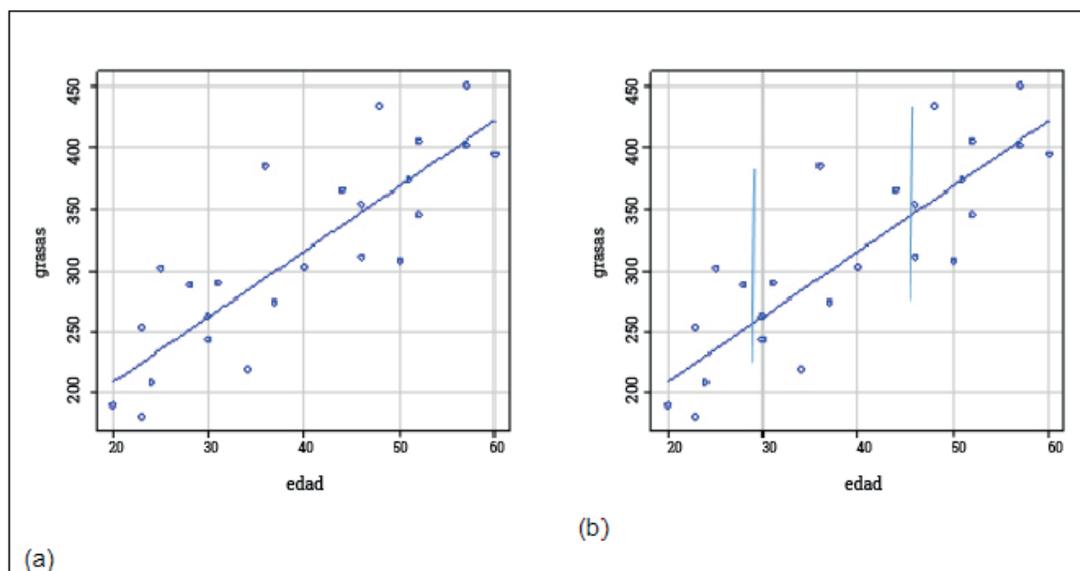


Figura 2: Ajuste de un modelo de regresión lineal simple, considerando edad-grasas

Observe la figura 2(a) donde se ajusta una línea recta que intenta pasar por el centro de la nube de puntos con el objetivo de modelar un comportamiento entre las variables edad y grasas. Ahora, observe la figura 2(b) donde se resaltan líneas verticales en las edades 30 y 45: note, por ejemplo, que existen 2 valores en el eje de las grasas (eje y) para la edad 30, es decir, las duplas (30, 263) y (30, 244). Esto, literalmente, indica que 2 individuos con la misma edad poseen cantidades de grasa en la sangre diferente. Lo anterior no sorprende en absoluto, pero si el objetivo es estimar un modelo lineal que relacione tales variables, entonces existe un efecto “perturbador” que ocasiona estas duplas y que es necesario minimizar.

Llevemos lo anterior a nuestro modelo: $grasas_{ij} = \alpha + \beta(edad)_i + \varepsilon_{ij}$, entonces la cantidad de grasa en la sangre cambia de forma constante (α, β) en la medida en que crece o decrece la $(edad)$, pero con un efecto aleatorio (ε_{ij}) causante de “perturbaciones” en los valores estimados. Es aquí donde entran en juego los tan atropellados supuestos básicos del modelo: $\varepsilon_{ij} \sim NID(0, \sigma^2)$. Teóricamente, $\varepsilon_{ij} = |y_{ij} - \hat{y}_{ij}|$ es decir, la diferencia entre el valor observado y el valor estimado por el modelo ajustado, entonces para hacer robusto el modelo se espera que tales efectos aleatorios, llamados comúnmente como errores, sigan una distribución normal e independiente con media cero y varianza constante. Pero, ¿cómo se traduce esto en nuestro el modelo?

Recordemos que el valor esperado de una variable aleatoria es el número que representa el valor medio del fenómeno que representa dicha variable, entonces lo que se desea es estimar el promedio de la cantidad de grasas para una edad particular, es decir:

$$\begin{aligned} \mu_{grasas/edad} = E(grasas_{ij}) &= E(\alpha + \beta(edad)_i + \varepsilon_{ij}) = \alpha + \beta(edad)_i + E(\varepsilon_{ij}) \\ &= \alpha + \beta(edad)_i \quad \text{sii} \quad \varepsilon_{ij} \sim NID(0, \sigma^2) \end{aligned}$$

Esto es: el modelo podrá estimar el valor esperado de la cantidad de grasas en función de la edad si y solo si $\varepsilon_{ij} \sim NID(0, \sigma^2)$ tal que $E(\varepsilon_{ij}) = 0$. En términos generales:

$$E(Y_{ij} / X = x_i) = \mu_{y/x} = \alpha + \beta x_i \quad \text{sii} \quad \varepsilon_{ij} \sim NID(0, \sigma^2)$$

Donde $\mu_{y/x}$ representa la esperanza de la distribución de Y correspondiente al valor dado de X .

Antes de continuar con nuestro estudio, es necesario recordar al lector que (matemáticamente) la diferencia entre las notaciones mayúsculas y minúsculas radica en la teoría de conjuntos. Es decir, cuando se menciona a X (en mayúscula) se refiere al dominio de la variable como el conjunto de todos los valores medidos (todas las edades en nuestro ejemplo), mientras que si se menciona a x (en minúscula) se refiere a un valor especificado en el conjunto, por ejemplo $x=edad=30$.

Teniendo en claro todo lo mencionado hasta ahora, es importante seguir explorando los elementos del modelo. Específicamente hablemos de los parámetros a estimar como medidas constantes en el modelo, es decir (α, β) . Como se puede apreciar en la figura 3, la ordenada al origen está representada por α , es decir, representa la dupla $(\alpha, x = 0)$. El nivel de inclinación de la recta estimada es representada por β , es decir, la pendiente de la recta, la cual denotará la tasa de cambio de y por cada incremento de x . Es ahora importante preguntarse: ¿Qué significa este contexto en nuestro ejemplo?

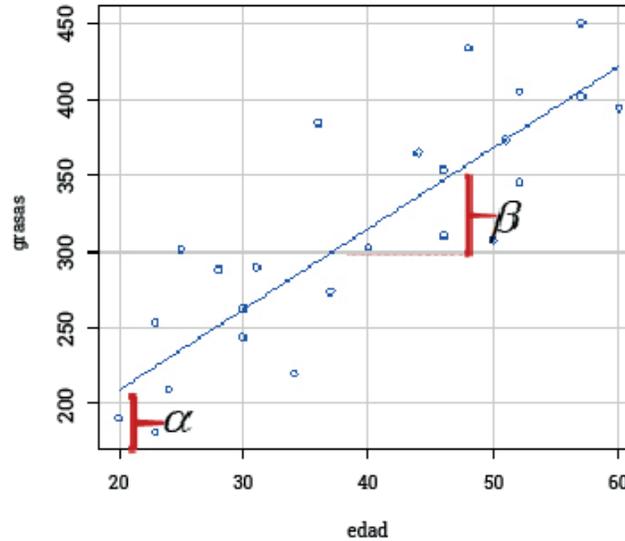


Figura 2: Parámetros del modelo de regresión lineal simple

Retomemos nuestro modelo:

$$E(\text{grasas}_{ij}) = \alpha + \beta(\text{edad})_i \quad \text{sii} \quad \varepsilon_{ij} \sim NID(0, \sigma^2)$$

Entonces, β el incremento en la cantidad de grasas a medida que cambia la edad. Es importante decir que por ser $\beta > 0$ hablo de incremento, pero pensar por favor, qué pasaría si $\beta < 0$. Por otro lado, indicará la cantidad de grasas en la sangre en la edad cero. Pero, **¿Tiene eso sentido real?**

Como estadístico lo que pretendo con este tipo de interrogantes es llamar la atención al lector en cuanto a la íntima relación que existe entre el modelo a estimar y el contexto real, razón por la cual la interdisciplinariedad en los análisis estadísticos es de vital importancia.

Continuando con los aspectos procedimentales del análisis de regresión, el siguiente paso debe ser la estimación de los parámetros mencionados (α, β) . Es decir, por medio de nuestros datos calcular los coeficientes de regresión muestral (estadísticos), que conllevará a un modelo estimado de la forma: $\hat{y}_{ijk} = a + bx_i$. El método más comúnmente usado, pero no el único, es el de Mínimos Cuadrados Ordinarios (MCO) donde el objetivo de estimación se centra en minimizar el efecto aleatorio (error) bajo el estricto supuesto de $\varepsilon_{ij} \sim NID(0, \sigma^2)$

Como se mencionó al inicio, el objetivo esta sesión es proporcionar al lector los cimientos básicos del análisis de regresión lineal, a efectos de propiciar un pensamiento crítico ante posibles estudios. Por tal razón, los cálculos inherentes en la estimación de los parámetros, la evaluación del modelo, pruebas de hipótesis, verificación de supuestos, entre otros, formarán parte de posteriores ediciones. Sin embargo, para tener una idea más clara hacia la tercera parte del presente escrito, es necesario que conversemos un poco del coeficiente de determinación.

¿Por qué es tan importante el coeficiente de determinación?

Como es lógico pensar, al ajustar un modelo de regresión lineal se pretende (aunque suene redundante) que la variabilidad estimada por la variable independiente sea superior al efecto aleatorio, pues solo así se podrá hablar del poder predictivo del modelo. Para tales efectos, es necesario el cálculo del coeficiente de determinación R^2 .

Matemáticamente, R^2 es un indicador porcentual de la variabilidad explicada del modelo en términos predictivos, por tanto se considera el cociente del cuadrado de la variabilidad conjunta (covarianza) y el producto de las respectivas varianzas, como se expresa en (3).

$$R^2 = \frac{\left(\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{Cov(XY)^2}{Var(X)Var(Y)} \quad (3)$$

En nuestro caso, un $R^2=0.80$ indicará que el 80% de la variabilidad de la cantidad de grasas en la sangre es explicada de forma lineal por la edad del individuo. Pero, ¡cuidado!, esto en ningún momento puede ser entendido como la adecuación del modelo o como una medida de la relación funcional entre las variables bajo estudio. El R^2 calculado ya está asumiendo una relación funcional, por lo que es tarea del investigador determinar si esa relación lineal es correcta en términos del fenómeno bajo estudio.

El coeficiente de determinación R^2 debe emplearse con mucho cuidado y, en principio, debe ser una medida referencial en términos de variabilidad. Es decir, valores bajos deben conllevar a medidas más robustas de evaluación del modelo como pruebas de hipótesis, bandas de confianza, pruebas de ajuste del modelo y (no menos importante) el contexto disciplinar del objetivo bajo estudio. Es un error común intentar aumentar el R^2 cuando lo correcto es ahondar en el fenómeno bajo estudio a fin de evaluar la relación funcional que estamos suponiendo. Y esto se debe precisar bien, pues muchas herramientas informáticas ofrecen la “evaluación del modelo” mediante un proceso obvio de eliminar/incorporar variables explicativas (siempre se puede aumentar el valor del añadiendo variables al

modelo), pero esto no siempre implica un mejor modelo, al contrario, se puede traducir en una falacia matemática si no se tiene suficiente conocimiento del fenómeno bajo estudio.

Como ya se ha mencionado, el objetivo del presente artículo es propiciar un pensamiento crítico al investigador ante puntos que comúnmente se usan y de los que a menudo se abusa en términos de regresión y correlación. Entonces, para avanzar, hago énfasis en que el R^2 sólo será válido en su capacidad predictiva cuando se ha concebido el modelo correctamente tanto en su aspecto funcional determinista como en su aspecto aleatorio.

Parte 2: Hablemos ahora de correlaciones

Para comenzar diré que, estadísticamente, la correlación mide el **grado de asociación** entre 2 variables que intervienen en una distribución bidimensional. Asociación no entendida como causa o dependencia, sino como una simple relación, tal vez justificada o tal vez producto del azar; pero siempre bajo unos criterios bien definidos como, por ejemplo, la normalidad conjunta (normalidad bivariada). Vamos poco a poco.

Retomemos nuestro ejemplo inicial. Recuerde que se desea estudiar la cantidad de grasas en la sangre en base a la edad y el peso de algunos individuos. Nutricionalmente se sabe que las variables peso y edad se pueden fijar para tal estudio, es decir, han sido predisuestas por el investigador en el contexto real. Ahora, lo que intentamos verificar es la asociación de tipo lineal de tales variables.

En efecto, el coeficiente de correlación lineal considera el cociente entre la covarianza de las variables bajo estudio, sobre la raíz cuadrada positiva del producto de las correspondientes varianzas. Esto es:

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (4)$$

La medida ρ hace referencia al parámetro, medida poblacional, y como es bien sabido los valores obtenidos de las muestras son estadísticos que fungen como estimadores. El estadístico más comúnmente usado para el parámetro ρ se muestra en la ecuación (5) y se conoce como el coeficiente de correlación **muestral** de Pearson.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5)$$

Note que las variables en cuestión se denotan en mayúsculas pues hacen referencia al conjunto de observaciones, en nuestro caso, observaciones pareadas de la forma $(x_i, y_i) = (\text{edad}_i, \text{grasas}_i)$. donde el subíndice corresponde al individuo observado.

Compare ahora las ecuaciones (3) y (4). Note que $r = \sqrt{R^2}$, y esto se debe a que R^2 no contiene información del signo de la pendiente, lo cual es fundamental para determinar la relación proporcional o inversamente proporcional de las variables. Pero, lo más resaltante aquí es que ambas medidas se refieren a relaciones de tipo lineal, es decir, obtener valores bajos solo implica la falta de relación de tipo lineal. Veamos esto en nuestro ejemplo.

La tabla 1 muestra la matriz de correlaciones de nuestras variables. Compare por favor los valores con lo observado en la figura (1) y note la relación lineal fuerte entre las variables edad y cantidad de grasas en la sangre, denotado como $r_{(\text{grasa}, \text{edad})} \approx 0.84$.

	Edad	Grasas	Peso
Edad	1	0.8374	0.2400
Grasas	0.8374	1	0.2652
Peso	0.2400	0.2652	1

Tabla 1: Correlaciones lineales edad-grasas y peso-grasas

El lector pudiera estar pensando, en términos nutricionales, que las variables peso-edad y peso-grasas están relacionadas. Entonces ¿por qué no se reflejan en la tabla? La razón es simple, observe de nuevo la figura 1(b) y note que la nube de puntos no es de la forma lineal $y=a+bx$. Es decir, no se duda en la relación existente entre las variables, aquí lo que verificamos es que esa asociación **no es de tipo lineal**.

En general, el coeficiente de correlación de Pearson, mide el grado de asociación lineal de una muestra aleatoria normal bivariada de tamaño n . Esta medida adimensional posee valores en el intervalo. Una medida cercana a los extremos indicará un grado de asociación lineal fuerte, donde los signos indican relación directamente proporcional en el caso positivo o indirectamente proporcional en el caso negativo.

Es estrictamente necesario verificar la significancia de tal medida, es decir, comprobar que la asociación no sea producto del azar o de un tamaño de muestra reducido. Por ello es vital mencionar que la ecuación (5) parte del supuesto de normalidad bivariada, es decir,

se debe verificar que la muestra cumpla de forma estricta tal consideración y, en casos del no cumplimiento o muestras reducidas, los coeficientes no paramétricos como el de Spearman o Kendall son los más adecuados. Ahora bien, si es verificado el supuesto de normalidad bivariada, entonces se debe proceder a pruebas de hipótesis para evaluar la significancia de las correlaciones.

Nuevamente hay que decir que los aspectos procedimentales de cálculo se escapan a los objetivos del presente escrito. Se invita al lector a revisar la amplia bibliografía existente en tales temas, en especial me atrevo a recomendar los clásicos (Hines & Montgomery, 1996) y (Walpole, Myers, Myers, & Ye, 2012), ambos con lenguaje simple y en castellano. Para aquellos un poco más atrevidos y con conocimientos de programación en R, recomiendo a (Heiberger & Holland, 2009).

Entonces, actualizados las bases inherentes a los análisis de regresión y correlación lineal, la tercera parte pretende mostrar los principales usos y abusos de estos grandes puntos.

Parte 3: Preconceptos y falsas ideas en los análisis de regresión y correlación

Para iniciar esta parte es necesario precisar que en los análisis de regresión, generalmente nos enfrentamos a un conjunto de datos que no se espera hayan sido tomados bajo condiciones estrictamente controladas y pocas veces se cuenta con repeticiones que conlleven a la estimación del error de tipo experimental. Entonces, informalmente, se puede decir que el objetivo aquí es extraer características no tan evidentes de la relación funcional entre dos variables, empleando técnicas estadísticas apropiadas y que conlleven a la estimación de una ecuación razonablemente válida. Teniendo esto en claro propongo al lector reflexionar sobre los dos puntos que siguen.

- La regresión no es una relación causa-efecto absoluta, solo puede asociarse un valor de y con uno de predicción x (relación funcional). Para aclarar lo anterior, sabemos que existe una relación obvia entre la altura y el peso de una persona. Entonces, por ejemplo, ¿puede modificar una persona su altura si altera su peso? O en nuestro caso de estudio: ¿es la edad la causa absoluta de la cantidad de grasas en la sangre? Si bien, sabemos la evidente relación entre tales variables, la única forma de determinar su relación causal es desarrollando un diseño experimental riguroso que permita controlar todas las variables, sumado al dominio y antecedentes científicos del objetivo bajo estudio. He aquí, nuevamente, la gran importancia de equipos interdisciplinarios en cualquier tipo de investigación

- No todo siempre es color “rosa” en los resultados y éste es un punto en el cual debo ser sumamente cruda. En muchos análisis estadísticos, medidas extremadamente buenas suelen estar enmascaradas por relaciones entre variables consideradas independientes o sesgos en la toma de datos. Si bien, el objetivo no es profundizar en fundamentos matemáticos, es necesario alertar al lector, por ejemplo, ante medidas de R^2 casi perfectas. Tal ejemplo implica grandes explicaciones; por ahora solo diré que ante tales casos es preferible no alegrarse y hacer una exhaustiva revisión del modelo propuesto y la muestra correspondiente en todo su ámbito (aleatorio y determinístico).

Continuando el esquema del presente escrito, precisemos un poco en el empleo desmesurado de las correlaciones. En términos experimentales, éste tipo de análisis se suele emplear cuando ninguna de las variables bajo estudio ha sido controlada y simplemente se desea saber si existe una posible **asociación** entre ellas. De hecho, una práctica común (no del todo errónea) es que se calculen primero medidas de correlación y posteriormente se proponga un modelo de regresión. Teniendo esto claro, nuevamente invito al lector a reflexionar sobre los siguientes puntos.

- Las correlaciones significativas estadísticamente expresan la existencia de una relación lineal entre dos variables y no implican causalidad. Es importante considerar que la regresión lineal explica una relación entre dos variables (x,y), donde una de ellas es dependiente de la otra. Sin embargo, esto no se aplica al concepto de correlación. La correlación mide el nivel de asociación lineal entre dos variables (x,y) pero sin asumir dependencia alguna, es decir, no implica “medida de cambio” de una variable con respecto a la otra. La correlación entre dos variables puede ser causada por: x influye sobre y, y influye sobre x o ambas variables están relacionadas con una tercera; es decir, en ningún momento se mide “causa”. **Causalidad siempre implica correlación, pero la correlación no siempre implica causalidad**
- La falacia ecológica: conocida como la atribución de características de una población a algunos individuos que forman parte de ella, sin considerar las características propias de tales individuos. Tal vez eso puede sonar un poco inocente, pero es uno de los principales problemas de no delimitar oportunamente la población y la muestra en el momento de elaborar conclusiones con validez estadística. Como un ejemplo clásico menciono el artículo: Chocolate consumption, cognitive function and nobel laureates (Messerli, 2012), en el que, mediante altas correlaciones entre el consumo per cápita de chocolate y el origen (nacionalidad) de los ganadores de premios Nobel, se concluye que el consumo de chocolate mejora la función cognitiva. Vale la pena mencionar que el mismo artículo menciona que no se poseía ningún conocimiento del consumo de chocolate de los laureados y que los resultados fueron simplemente extrapolados. Lo más resaltante de todo es que, a pesar de las altas críticas recibidas, dicho artículo aún está disponible en la web. He aquí una razón más de por qué considero la alfabetización estadística como una necesidad irrenunciable ante cualquier tipo de estudio

- Como se puede observar en la ecuación (5), las medidas provenientes del cálculo de correlaciones no son números matemáticamente simples, por tanto, no se pueden tratar con medidas aritméticas habituales. Me explico mejor: mi experiencia con las ciencias biológicas me ha permitido entender la importancia de, por ejemplo, la caracterización de especies mediante correlaciones de variables. Entonces, como parte de un análisis exploratorio, podríamos pensar en seleccionar individuos en diferentes hábitats⁵ y determinar las medidas de correlaciones promedio de las variables teóricamente consideradas, estudio que podría plantearse perfectamente en algunos análisis de tipo social. Ahora bien, para determinar las correlaciones promedio es necesario realizar transformaciones a tales medidas (métodos paramétricos) o emplear índices de Spearman (métodos no paramétricos). Nuevamente, con el ánimo de no detallar procedimientos matemáticos en el tema, recomiendo un texto de explicación bien sencilla titulado: *Measuring Behaviour: An introductory guide*. (Martin & Batenson, 2007).

Parte 4: Conclusiones y/o recomendaciones

El presente escrito está dirigido a no estadísticos, dado que el objetivo central es crear ruido en el lector, es decir, refrescar o rescatar algunos cimientos básicos que poseen los análisis de regresión y correlación, comenzando con los aspectos lineales por su gran aplicabilidad. A este punto es al que llamo **alfabetización estadística** como una necesidad irrenunciable: considero importante que manejemos un mismo idioma a efectos de poder crear grupos interdisciplinarios eficientes y más aún, que cualquier profesional pueda leer de forma crítica cualquier publicación referencial. Quiero hacer especial énfasis en lo último, pues actualmente poseemos un exceso de información y, como profesionales, debemos discriminar entre revistas indexadas serias en sus publicaciones donde los resultados no solo sean coherentes sino que posean el rigor teórico y procedimental necesario.

Actualmente existe una gran variedad de “paquetes informáticos” que prometen mil y un procedimientos. El papel de la Estadística no es simplemente procesar, se tiene una gran responsabilidad teórica en la fase del estudio, es decir, desde su concepción hipotética hasta el análisis de los resultados. Para no aburrir mucho al lector en este punto, quiero finalizar parafraseando las palabras atribuidas a R. Fisher (1890-1962), en las que expresa que “llamar a un estadístico después de que el experimento ya está hecho, es llamarle para que realice un examen post-mortem: solamente podrá decir de que murió el experimento”. Lo anterior no es más que una recomendación a la planificación exhaustiva en cualquier estudio, es decir, tener en claro en primer lugar las hipótesis que se desean probar, no simplemente recoger datos y ver qué sale y recordar que “Todos los modelos son falsos, pero algunos son útiles” (G. Box, 1919-2013).

⁵ Recordar que un hábitat se concibe como el espacio que reúne condiciones/características físicas y biológicas necesarias para la reproducción y supervivencia de una especie.

REFERENCIAS BIBLIOGRÁFICAS

- Costa S., E., Villa-Murillo, A., Ramirez, C., Gaete Sierra, J., González-Romero, F., & Reyes-Lilo, D. (2001). La investigación científica desde las metodologías cuantitativa, cualitativa, mixta y sus aspectos éticos. Viña del Mar, Chile: Universidad Viña del Mar. Obtenido de <https://hdl.handle.net/20.500.12536/1210>
- Heiberger, R., & Holland, B. (2009). Statistical analysis and data display. An intermediate course with examples in R. Springer.
- Hines, W., & Montgomery, D. (1996). Probabilidad y estadística para ingeniería y administración. Compañía editorial continental S.A.
- Martin, P., & Batenson, P. (2007). Measuring Behaviour: An introductory guide (Third edition ed.). Cambridge University Press.
- Messerli, F. (2012). Chocolate consumptions, cognitive function and novel laurates. The new England journal of medicine.
- Walpole, R., Myers, R., Myers, S., & Ye, K. (2012). Probabilidad y estadística para ingeniería y ciencias. Pearson Educación.

