



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Construcción, Análisis e Interpretación de Gráficos Estadísticos con estudiantes de ciclo IV mediante el trabajo por proyectos

Hernán Augusto Rojas Sanabria

Universidad Nacional de Colombia
Facultad de Ciencias
Maestría en Enseñanza de las Ciencias Exactas y Naturales
Bogotá, Colombia
Año 2017

Construcción, Análisis e Interpretación de Gráficos Estadísticos con estudiantes de ciclo IV mediante el trabajo por proyectos

Hernán Augusto Rojas Sanabria

Trabajo de investigación presentado como requisito parcial para optar al título de:
Magister en Enseñanza de las Ciencias Exactas y Naturales

Director:

Dr. Luis Fernando Grajales Hernández

Universidad Nacional de Colombia

Facultad de Ciencias

Bogotá, Colombia

Año 2017

A mis queridos padres: Rafael Rojas y Myriam Sanabria

Agradecimientos

Expreso mi agradecimiento al profesor Luis Fernando Grajales Hernández por su continua atención, colaboración, asesoría y dirección en la realización de este trabajo, su labor y conocimiento me permitió aprender pautas acerca de la investigación en ciencias y desarrollar un vínculo más fuerte con la estadística.

Además, deseo reconocer y agradecer la oportuna atención y dirección recibida de la profesora Myriam Margarita Acevedo Caicedo al iniciar el desarrollo de mi propuesta de grado. Su orientación me permitió desarrollar habilidades para la investigación en educación matemática a partir de mi labor en el aula de clase.

Por otra parte, agradezco a mis estudiantes de noveno grado del colegio Juan Evangelista Gómez por su esfuerzo, dedicación y empeño en la realización de las actividades académicas propuestas en este trabajo.

Resumen

En este trabajo se elabora una estrategia didáctica por medio del proyecto *Estadísticas de sexualidad*, utilizando datos referentes a planificación familiar y algunas características generales de las mujeres en edad fértil, para la construcción, interpretación y análisis de gráficos estadísticos. Este proyecto se implementa con 28 estudiantes de un grado noveno y se analizan los gráficos producidos por medio del diseño de niveles de respuesta teniendo en cuenta los elementos estructurales de un gráfico estadístico y la comprensión de información estadística. Los resultados muestran que la mayoría de estudiantes identifican las variables y los datos a representar en la construcción de un gráfico estadístico, pero se presentan algunos errores con respecto a la utilización de elementos estructurales como: un título, etiquetas o algunos convenios de construcción.

Palabras clave: Gráficos estadísticos, trabajo con proyectos, software R.

Abstract

In this work a didactic strategy is elaborated through the Sexuality Statistics project, using data referring to family planning and some general characteristics of women of childbearing age, for the construction, interpretation and analysis of statistical graphs. This project is implemented with 28 students of a ninth grade and the graphs produced are analyzed through the design of response levels taking into account the structural elements of a statistical graph and the understanding of statistical information. The results show that the majority of students identify the variables and the data to be represented in the construction of a statistical graph, but some errors are presented regarding the use of structural elements such as: a title, labels or some construction agreements.

Keywords: Statistical graphics, work with projects, R software.

Contenido

	<u>Pág.</u>
Resumen.....	V
Lista de figuras	VIII
Lista de tablas	IX
Lista de Símbolos y abreviaturas	XI
Introducción	1
1. Capítulo 1: Contextualización	3
1.1. Ubicación geográfica y contexto educativo.....	3
1.2. Planteamiento del problema.....	3
1.3. Pregunta de investigación	4
1.4. Objetivos	5
1.4.1. General.....	5
1.4.2. Específicos	5
2. Capítulo 2: Referentes Teóricos	6
2.1. Marco Histórico y Epistemológico.....	6
2.2. Marco Disciplinar: Estadística Descriptiva	8
2.2.1. Variables Estadísticas	8
2.2.2. Tablas Estadísticas	10
2.2.2.1. Tablas de contingencia para dos variables cualitativas	12
2.2.2.2. Tablas de frecuencias para una variable cuantitativa	13
2.2.3. Gráficos Estadísticos	14
2.2.3.1. Gráficos para variables cualitativas	14
2.2.3.2. Gráficos para variables cuantitativas	15
2.3. Marco Didáctico: Trabajo con Proyectos	19
2.4. Marco Metodológico: Investigación Acción.....	20
3. Capítulo 3: Estrategia didáctica	21
3.1. Prueba diagnóstica	21
3.2. El proyecto: Estadísticas de sexualidad	24
3.2.1. Objetivos del proyecto	24
3.2.2. Los datos.....	24
3.2.3. La actividad	25
3.3. Utilización de un software estadístico.....	38
3.3.1. Utilización de comandos básicos en R-Project	38
3.3.2. Exploración de Excel para la elaboración de un boxplot	39
4. Capítulo 4: Estrategia didáctica: Resultados y análisis	41
4.1. Prueba diagnóstica: análisis de resultados.....	41
4.1.1. Análisis de las preguntas de opción múltiple	41
4.1.2. Análisis estadístico inferencial de las pruebas: diagnóstica (antes) y final (después)	42
4.1.3. Análisis cualitativo de las preguntas abiertas.....	46

4.2. Proyecto Estadísticas de sexualidad: análisis de resultados	53
5. Conclusiones y recomendaciones	62
5.1 Conclusiones sobre los objetivos	62
5.2 Recomendaciones	66
A. Anexo: Prueba diagnóstica	68
B. Anexo: Hoja de códigos. Estadísticas de sexualidad	72
C. Anexo: Proyecto. Estadísticas de sexualidad.....	74
D. Anexo: Actividad de reconocimiento en R-Project.....	75
E. Anexo: Importar datos a R-Project	78
F. Anexo: Modificaciones a la prueba diagnóstica	80
Apéndice.....	82
Bibliografía	86

Lista de figuras

	<u>Pág.</u>
<u>Figura 1:</u> Ejemplos de boxplots	18
<u>Figura 2:</u> Gráfico de Pareto para uso de métodos anticonceptivos	26
<u>Figura 3:</u> Reacción si se descubre que un hijo es homosexual según el sexo biológico de la cabeza de hogar	28
<u>Figura 4:</u> Histograma para las alturas	30
<u>Figura 5:</u> Boxplot para las alturas	31
<u>Figura 6:</u> Histograma para los pesos	32
<u>Figura 7:</u> Diagrama de barras, en porcentaje, para el conocimiento del periodo fértil en el ciclo ovulatorio	34
<u>Figura 8:</u> Diagrama de sectores para el conocimiento del periodo fértil en el ciclo ovulatorio	34
<u>Figura 9:</u> Diagrama de barras para el conocimiento, en porcentaje, del periodo fértil en el ciclo ovulatorio entre las mujeres que pertenecen al nivel educativo secundaria	36
<u>Figura 10:</u> Boxplots del número de hijos nacidos según los niveles educativos: primaria, secundaria y mayor	37
<u>Figura 11:</u> Macro en Excel para elaborar un boxplot	39
<u>Figura 12:</u> Dos boxplot construidos con diferente software para un mismo conjunto de datos	40
<u>Figura 13:</u> Ejemplo de un gráfico nivel 1 producido por un estudiante	47
<u>Figura 14:</u> Ejemplo de un gráfico nivel 2 producido por un estudiante	48
<u>Figura 15:</u> Ejemplo de un gráfico nivel 3 producido por un estudiante	48
<u>Figura 16:</u> Ejemplos de gráficos producidos por estudiantes según el nivel de respuesta	49
<u>Figura 17:</u> Gráficos producidos por un estudiante para responder la pregunta 9 de la prueba en los dos momentos de aplicación	51
<u>Figura 18:</u> Gráficos producidos por un estudiante para responder la pregunta 10 de la prueba en los dos momentos de aplicación	52
<u>Figura 19:</u> Algunos errores en los gráficos producidos por estudiantes para el desarrollo de la primera pregunta	56
<u>Figura 20:</u> Ejemplo de un gráfico nivel 4 para la primera pregunta	56
<u>Figura 21:</u> Ejemplo de un gráfico nivel 3 para la segunda pregunta	57
<u>Figura 22:</u> Ejemplo de un gráfico nivel 3 para la tercera pregunta	58
<u>Figura 23:</u> Ejemplos de algunos errores encontrados en la construcción de un histograma	59
<u>Figura 24:</u> Histograma para las alturas en R	82
<u>Figura 25:</u> Boxplot para las alturas en R	84
<u>Figura 26:</u> Histograma para los pesos en R	84

Lista de tablas

	<u>Pág.</u>
<u>Tabla 1:</u> Formato de una tabla de distribución de frecuencias	11
<u>Tabla 2:</u> Formato general de una tabla de contingencia IxJ	12
<u>Tabla 3:</u> Análisis a priori de la prueba diagnóstica	21
<u>Tabla 4:</u> Datos de planificación familiar y percepción de la educación sexual	25
<u>Tabla 5:</u> Número de mujeres por cada método anticonceptivo	26
<u>Tabla 6:</u> Convenciones usadas respecto a lo que “haría si se descubre que uno de sus hijos es homosexual”	27
<u>Tabla 7:</u> Clasificación de mujeres según el sexo biológico de la cabeza de hogar y la reacción si se descubre que un hijo es homosexual	28
<u>Tabla 8:</u> Distribución de frecuencias para las alturas	29
<u>Tabla 9:</u> Distribución de frecuencias para los pesos	32
<u>Tabla 10:</u> Frecuencias y porcentajes para el conocimiento del periodo fértil en el ciclo ovulatorio	33
<u>Tabla 11:</u> Clasificación de mujeres según el conocimiento del periodo fértil en el ciclo ovulatorio y el nivel educativo	35
<u>Tabla 12:</u> Frecuencias para el conocimiento del periodo fértil en el ciclo ovulatorio entre las mujeres con nivel educativo secundaria	35
<u>Tabla 13:</u> Clasificación de mujeres según el número de hijos nacidos y el nivel educativo	36
<u>Tabla 14:</u> Media y mediana del número de hijos nacidos según el nivel educativo	37
<u>Tabla 15:</u> Elementos para la elaboración de boxplots del número de hijos nacidos según los niveles educativos primaria, secundaria y mayor	38
<u>Tabla 16:</u> Resultados de la prueba diagnóstica	41
<u>Tabla 17:</u> Resultados de la prueba final	42
<u>Tabla 18:</u> Tabulación de datos para el test de McNemar	43
<u>Tabla 19:</u> Datos de la prueba: pregunta 1	44
<u>Tabla 20:</u> Datos de la prueba: pregunta 3	44
<u>Tabla 21:</u> Datos de la prueba: pregunta 4	44
<u>Tabla 22:</u> Datos de la prueba: pregunta 5	45
<u>Tabla 23:</u> Datos de la prueba: pregunta 6	45
<u>Tabla 24:</u> Datos de la prueba: pregunta 7	45
<u>Tabla 25:</u> Estudiantes que utilizaron elementos estructurales en sus gráficos durante la prueba diagnóstica	46
<u>Tabla 26:</u> Clasificación de estudiantes según el nivel de respuesta durante la prueba diagnóstica	50
<u>Tabla 27:</u> Estudiantes que utilizaron elementos estructurales en sus gráficos durante la prueba final	50
<u>Tabla 28:</u> Clasificación de estudiantes según el nivel de respuesta durante la prueba final	52

<u>Tabla 29:</u>	<u>Información acerca de los gráficos producidos por los estudiantes para las primeras tres preguntas del proyecto</u>	54
<u>Tabla 30:</u>	<u>Estudiantes que utilizaron elementos estructurales en los gráficos presentados en el proyecto</u>	54
<u>Tabla 31:</u>	<u>Clasificación de estudiantes según el nivel de respuesta en la construcción de gráficos estadísticos</u>	55
<u>Tabla 32:</u>	<u>Algunos comandos de R-Project utilizados en estadística descriptiva</u>	75
<u>Tabla 33:</u>	<u>Distribución de frecuencias para las alturas en R</u>	83
<u>Tabla 34:</u>	<u>Distribución de frecuencias para los pesos en R</u>	85

Lista de Símbolos y abreviaturas

Símbolos con letras latinas

Símbolo	Término
f	Frecuencia absoluta
F	Frecuencia absoluta acumulada
h	Frecuencia relativa
H	Frecuencia relativa acumulada
$\log_{10}(x)$	Logaritmo en base 10 de x
R_x	Rango observado de una variable aleatoria
$x_{mín}$	Valor mínimo de un conjunto de datos
$x_{máx}$	Valor máximo de un conjunto de datos

Abreviaturas

Abreviatura	Término
<i>ETS</i>	Enfermedades Transmisibles Sexualmente
<i>ICFES</i>	Instituto Colombiano para el Fomento de la Educación Superior
<i>MEN</i>	Ministerio de Educación Nacional

Introducción

En Colombia, las directrices curriculares planteadas desde el Ministerio de Educación Nacional por medio de los Estándares Básicos de Competencias en matemáticas y los Derechos Básicos de Aprendizaje presentan procesos con respecto al desarrollo de competencias que se dividen en cinco tipos de pensamiento: el numérico, el espacial, el métrico, el aleatorio o probabilístico y el variacional. Específicamente, en el pensamiento aleatorio y los sistemas de datos se utilizan estrategias como la exploración de datos siendo importante que el estudiante interprete, compare y analice resultados publicados en medios de comunicación (MEN, 2006).

Además, existe la necesidad de formar ciudadanos capaces de tratar con diversos tipos de informaciones estadísticas e interpretar representaciones en distintos contextos de su vida como revistas, programas de televisión o evaluaciones (Arteaga, Batanero, Cañadas y Contreras, 2011). No obstante, desde mi experiencia en el aula he identificado la dificultad de algunos estudiantes al momento de interpretar los datos de un gráfico estadístico y comprender sus elementos estructurales.

De manera que, en la búsqueda de una estrategia didáctica que desarrolle mayores niveles de comprensión, interpretación y análisis de gráficos estadísticos, se implementa en este documento el *trabajo con proyectos* como metodología de enseñanza para abordar, desde la estadística descriptiva, un conjunto de datos referentes a temas de planificación y percepción de la sexualidad -tema que despierta el interés de jóvenes y adolescentes-.

El desarrollo de este documento se realiza en cinco capítulos. En el primer capítulo, se comienza con una breve descripción del contexto y la ubicación geográfica donde se lleva a cabo la investigación, seguido por el planteamiento del problema teniendo en cuenta algunas directrices curriculares en la formación estadística y algunas dificultades en la comprensión de gráficos estadísticos. Finalmente, se plantea la pregunta de investigación y los objetivos propuestos.

En el segundo capítulo se elaboran los marcos histórico y epistemológico, disciplinar, didáctico y metodológico que se tomaron como referencia para esta investigación. Específicamente, en el marco histórico y epistemológico se realiza una síntesis acerca del surgimiento de los gráficos estadísticos y de algunos niveles para su comprensión. Posteriormente, se desarrolla el marco disciplinar con base en la estadística descriptiva y

con énfasis en la construcción de gráficos estadísticos. Finalmente, se aborda *El trabajo con proyectos* como metodología de enseñanza y el tipo de investigación que se utiliza desde los marcos didáctico y metodológico, respectivamente.

En el tercer capítulo se describe tanto la prueba diagnóstica utilizada para identificar los conocimientos previos de los estudiantes como la estrategia didáctica diseñada por medio del proyecto de *Estadísticas de sexualidad*, realizando un análisis y una propuesta de solución para sus preguntas. Adicionalmente, se mencionan algunos detalles con respecto a la exploración y utilización de un software estadístico (como R-Project o Excel) para el análisis de un conjunto de datos.

En el capítulo cuatro se analizan los datos recolectados de la prueba diagnóstica, antes y después de aplicarse el proyecto, y se realiza un análisis estadístico inferencial utilizando el test de McNemar (teniendo en cuenta que se tienen los datos del mismo estudiante, antes y después de la intervención) para el caso de las preguntas de selección múltiple. Luego, con base en los informes elaborados por los estudiantes para responder a los proyectos, se realiza un análisis cualitativo con énfasis en la construcción de los gráficos estadísticos teniendo en cuenta algunos niveles de respuesta elaborados en este trabajo.

Finalmente, en el capítulo cinco se elaboran las conclusiones de la investigación y las recomendaciones para aquellos educadores que desean poner en práctica la estrategia didáctica utilizada en este trabajo.

Capítulo 1: Contextualización

Este capítulo aborda aspectos del contexto donde se ejecuta la propuesta de investigación teniendo en cuenta la ubicación geográfica, el planteamiento del problema, algunas directrices curriculares, algunas dificultades en la construcción, comprensión e interpretación de gráficos estadísticos, la pregunta de investigación y los objetivos propuestos.

1.1. Ubicación geográfica y contexto educativo

El contexto de esta propuesta de investigación se enmarca en la Institución Educativa Distrital Juan Evangelista Gómez ubicado en el barrio La Victoria de la localidad cuarta de San Cristóbal (al sur oriente de Bogotá), donde se trabaja el PEI *Construyendo el saber con equidad y autonomía*, por medio del modelo pedagógico socio crítico.

El colegio se caracteriza por tener una población estudiantil de estratos uno y dos, con algunas problemáticas sociales como: pertenecer a una pandilla, consumo de sustancias psicoactivas o embarazos en adolescentes. Para esta propuesta, se trabaja (inicialmente) con 29 estudiantes de ciclo IV -grado noveno- pertenecientes a la jornada tarde, con la siguiente distribución etaria: 48.28% de 14 años, 17.24% de 15 años, 17.24% de 16 años y 17.24% de 17 años.

1.2. Planteamiento del problema

Para esta investigación se aborda un tema de la estadística porque se considera una disciplina necesaria para el desarrollo de la sociedad por sus múltiples aplicaciones en las diversas áreas del saber –naturaleza interdisciplinaria-, de manera que su enseñanza resulta importante por la necesidad de generar una cultura estadística en el ciudadano, que le permita participar, interpretar y evaluar críticamente información. En consecuencia, una persona culta debe tratar con diversos tipos de información estadística y sus representaciones, detectando los posibles errores que puedan distorsionar la información presentada (Arteaga et al., 2011).

La importancia de la formación estadística del ciudadano se resalta también en nuestro país desde las directrices curriculares del MEN (Lineamientos y Estándares Básicos) y desde

los énfasis de la evaluación externa. ICFES (2016) considera dentro de las competencias matemáticas a evaluar, la referida a la comprensión y transformación de información que aparece en tablas, gráficos, diagramas, entre otros, para extraer de ellos información relevante que permita establecer relaciones matemáticas y encontrar tendencias y patrones. Más específicamente, uno de los desempeños que evalúa el ICFES en la prueba Saber Noveno es “resolver y formular problemas en diferentes contextos, que requieran hacer inferencias a partir de un conjunto de datos estadísticos provenientes de diferentes fuentes” (ICFES, 2015, p.7).

Sin embargo, (con mi experiencia docente) se evidencia la dificultad que tienen los estudiantes de ciclo IV –grados octavo y noveno- al momento de construir gráficos estadísticos, comprender e interpretar la información que reciben de diversas fuentes y adoptar una postura crítica y analítica, respecto, por ejemplo, a datos provenientes de medios de comunicación o de situaciones problema propuestas en las evaluaciones.

De acuerdo con Arteaga et al. (2011), la dificultad anterior se puede originar en que los gráficos, un potente instrumento para comunicar y resumir información, son objetos semióticos complejos que poseen elementos estructurales (con sus propios convenios de construcción e interpretación) como títulos, etiquetas, ejes, marcas de referencia, escalas, especificadores, entre otros. Adicionalmente, es posible que la comprensión de los gráficos estadísticos se dificulte cuando hay errores en su construcción, por ejemplo, en la representación de escalas numéricas en la recta real, la representación de intervalos, por falta de conocimiento de proporcionalidad y porcentaje, entre otros.

1.3. Pregunta de investigación

De la problemática antes mencionada surge la siguiente pregunta de investigación:

¿Qué tipo de estrategia didáctica es adecuada para trabajar con los estudiantes del ciclo IV, la construcción, lectura, interpretación y análisis de gráficos estadísticos provenientes de diversos contextos?

Para responder a la anterior pregunta planteada de esta propuesta de trabajo, se pretende desarrollar una estrategia didáctica en la que se utilice la metodología del trabajo por proyectos, de tal forma que los estudiantes investiguen sobre situaciones, preguntas y problemas del contexto con relación a sus propios intereses, representen gráficamente y

analicen información numérica, estudien la relación entre variables –bajo la supervisión y orientación del docente-, y luego contrasten con gráficos elaborados por Microsoft Excel o R-Project (software libre).

Además, se pretenden establecer conexiones entre la escuela y la vida cotidiana por medio de lectura crítica de gráficos utilizados en medios de comunicación y en diversos contextos de su vida, de tal forma que puedan adquirir habilidades para no sólo leer e interpretar datos, sino valorar críticamente, hasta donde sea posible, el método de recolección de la información.

1.4. Objetivos

1.4.1. Objetivo General

Desarrollar una estrategia didáctica para trabajar con los estudiantes del ciclo IV, la construcción, lectura, interpretación y análisis de gráficos estadísticos mediante la metodología del trabajo por proyectos.

1.4.2. Objetivos específicos

- Identificar conocimientos previos de los estudiantes respecto a la construcción, lectura, análisis y uso de información que aparece en gráficos estadísticos.
- Determinar aspectos epistemológicos, disciplinares y didácticos relacionados con la lectura, interpretación y análisis de gráficos estadísticos.
- Seleccionar situaciones, contextos e información pertinente para estructurar proyectos y diseñar actividades de la estrategia didáctica.
- Proponer ejercicios que conduzcan al diseño de macros simples en la hoja de cálculo de Excel para construir gráficos sencillos (opcionalmente, se enseñará la construcción de algunos gráficos en R-Project [software estadístico gratis]).
- Estructurar la estrategia didáctica mediante el trabajo por proyectos.
- Implementar y evaluar la estrategia didáctica.

Capítulo 2: Referentes Teóricos

Considerando que los gráficos estadísticos son el objeto de estudio de este trabajo, en este capítulo se elaboran los marcos histórico y epistemológico, disciplinar, didáctico y metodológico que sirven de sustento teórico para esta investigación. En primer lugar, se tratan, en el marco histórico y epistemológico, aspectos relevantes de su evolución histórica, dificultades en su comprensión y niveles de complejidad propuestos por diversos autores. Luego, en el marco disciplinar se aborda la estadística descriptiva, realizando una síntesis de algunos conceptos preliminares para la construcción de gráficos estadísticos y de aquellos gráficos que se utilizan en la básica secundaria.

Además, en el marco didáctico se realiza una breve descripción del Trabajo con proyectos como metodología de enseñanza y, finalmente, se describe la metodología de investigación acción utilizada en este trabajo por medio de las diversas fases en que se desarrolla.

2.1. Marco Histórico y Epistemológico

La síntesis de la evolución histórica y epistemológica de gráficos estadísticos se redacta utilizando los textos *Guía para la presentación de gráficos estadísticos* (INEI, 2009) y *Análisis de gráficos estadísticos elaborados en un proyecto de análisis de datos* (Arteaga, 2008).

A partir de la revolución científica del siglo XVII, los gráficos estadísticos surgen para ilustrar la información utilizando fundamentos del sistema de coordenadas cartesianas (creado por René Descartes), los cuales se modificaron parcialmente por Michael Van Langren para ilustrar las variaciones en la determinación de la longitud entre dos ciudades. La utilización de estos gráficos estadísticos aumenta en los siglos XVII y XVIII, consiguiendo su apogeo cuando William Playfair, político y economista inglés, expone su utilidad manifestando que permiten una comunicación más eficiente que las tablas de frecuencia e introduce, en el año 1786, 43 gráficos de series de tiempo y el gráfico de barras.

A pesar de que Playfair es el inventor de los gráficos lineales, de barras y de sectores, un lenguaje visual que tiene aplicaciones en las ciencias y en el comercio, sus hallazgos no son entendidos por sus contemporáneos. Solamente hasta el siglo XIX, se utilizan de nuevo

estas herramientas de análisis de datos y progresan a tal punto que se construyen una combinación de diferentes tipos de ellos para mostrar datos multidimensionales.

En el siglo XX, los gráficos estadísticos acompañaron la evolución de la ciencia, la estadística y la tecnología. Incluso, en el año 1924 el sociólogo austriaco Otto Neurath forma el Sistema Internacional de Educación Pictórico Tipográfica, lo que se puede considerar como la primera aproximación práctica a la enseñanza de este lenguaje visual.

Desde el punto de vista educativo, la lectura o comprensión de gráficos estadísticos a partir de un conjunto de datos puede resultar un proceso complejo que depende del tipo de datos o del tamaño y variación de éstos (Friel, Curcio y Bright, 2001). Es por esto que al surgir nuevos gráficos más complejos y elaborados se necesita de una mayor capacidad de recepción y análisis para su comprensión.

Estas dificultades en la comprensión dieron lugar a obstáculos en la lectura, interpretación y análisis de datos, los cuales fueron estudiados por Bertin (1967) para luego plantear niveles en la comprensión: la extracción de datos (la relación entre ejes y la lectura de coordenadas), la extracción de tendencias (Capacidad de comparar subconjuntos de datos desde el gráfico) y el análisis de la estructura de los datos (identificando las tendencias y agrupaciones).

Para complementar el trabajo de Bertin, Curcio (1989) redefine estos tres niveles como “leer entre los datos” (lectura del gráfico sin interpretación), “leer dentro de los datos” (interpretación de la relación entre datos) y “leer más allá de los datos” (capacidad de realizar inferencias y predicciones de datos). Además, este autor muestra que a medida que aumentan los niveles, simultáneo a las edades de los estudiantes, también aumentan las dificultades.

Sin embargo, para Gerber, Boulton-Lewis y Bruce (1995) la comprensión de información estadística se puede estudiar desde siete niveles de comprensión de gráficos que se presentan a continuación:

- Nivel 1. No se asocia la información de los datos con la realidad.
- Niveles 2 y 3. Hay una interpretación parcial de los datos, de manera que no se aprecia el propósito del gráfico (nivel 2) o, aunque se aprecia el propósito del gráfico, no se consiguen conclusiones globales (nivel 3).

- Niveles 4, 5 y 6. Aunque se llega a una síntesis global, puede sólo limitarse al análisis de una sola variable (nivel 4), análisis de varias variables (nivel 5) o apoyar o refutar sus hipótesis (nivel 6).
- Nivel 7. Hay extrapolaciones y predicciones de datos.

Es por esto que -desde el punto de vista didáctico- algunos autores como Pedro Arteaga, Carmen Batanero, Gustavo Cañadas y Miguel Contreras han considerado a los gráficos como objetos semióticos complejos compuestos por elementos como título, ejes, escalas, marcas del eje y elementos geométricos propios del gráfico, utilizando diversos tipos de lenguaje como verbal, gráfico y simbólico, en donde hay una correspondencia (establecida por el estudiante) entre un antecedente y un consecuente para leer o interpretar la información.

Esta complejidad puede llevar al estudiante a cometer errores en la construcción, lectura e interpretación de gráficos estadísticos. Por ejemplo, en los niveles superiores de la escuela es necesario tener conocimientos de proporción y porcentaje para una correcta interpretación de gráficos de barras (Arteaga et al., 2011), de modo que, si un estudiante tiene errores en estos conceptos preliminares, la correspondencia que realiza del gráfico le genera una construcción, lectura e interpretación errónea.

2.2. Marco Disciplinar: Estadística Descriptiva

Para el desarrollo y síntesis de este marco disciplinar se utilizan los textos *Análisis de tablas de contingencia bidimensionales* (Aguilera, 2005), *Estadística descriptiva y probabilidades* (Sosa, J., Ospina, L. y Berdugo, E., 2013, pp. 17-65), *Escalas de medición* (Coronado, J., 2007, pp. 104 -125) y entrevistas con el profesor Luis F. Grajales H. (2016).

2.2.1. Variables Estadísticas

En una investigación estadística es necesario delimitar la cuestión a investigar para reconocer las variables preponderantes del estudio.

“Una **variable** es una característica observable o medible de un individuo, que se describe acorde a una escala de medición bien definida. Cada rasgo, aspecto o característica de una población constituye una variable” (Sosa et al., 2013, p. 7).

Las variables se pueden clasificar en cualitativas o cuantitativas.

Las **variables cualitativas** se expresan como categorías o atributos que se miden por medio de las **escalas categóricas nominal u ordinal**.

En la **escala nominal** se clasifican las unidades de estudio en categorías teniendo en cuenta una o más características, atributos o propiedades observadas. A cada categoría le corresponde un nombre que puede ser alfabético o alfanumérico, para representar una etiqueta o identificador. Es necesario resaltar que en esta categoría los números no se pueden manipular aritméticamente puesto que cumplen una función puramente de asignación. Adicionalmente, las variables nominales se pueden clasificar en **dicotómicas** cuando incluyen únicamente dos categorías, o **politómicas** cuando contiene tres o más categorías.

En la **escala ordinal** las observaciones pueden colocarse en un orden relativo respecto a la característica que se evalúa, por lo que las características de los datos se clasifican u ordenan teniendo en cuenta una característica especial que poseen. Una vez los objetos están clasificados por una característica, es posible determinar qué objeto tiene más o menos de una característica en comparación con otro, pero no se puede decir en cuánto difieren. A diferencia de la escala nominal, las etiquetas de las categorías indican una jerarquía.

Tanto en la escala nominal como en la ordinal sólo se permite la identificación de la moda (la categoría con mayor frecuencia) y el cálculo e interpretación de proporciones, porcentajes o razones.

Las **variables cuantitativas** se registran en forma numérica y expresan una cantidad. Estas variables se clasifican en discretas y continuas.

Las **variables discretas** son aquellas que no admiten un valor intermedio entre dos valores cualesquiera de la variable.

Las **variables continuas** son aquellas que siempre admiten un valor intermedio entre dos valores cualesquiera de la variable.

Para la medición de variables cuantitativas se utilizan las escalas numéricas por **intervalo o razón**.

En la **escala de intervalo**, los elementos se ordenan por la magnitud del atributo que representan y se utilizan intervalos iguales entre las unidades de medida.

Además, es posible asignar un número a cada objeto de tal forma que la diferencia entre los objetos se refleje por la diferencia entre los números asignados. En este tipo de medida se puede utilizar cualquier unidad y el cero se establece, según Sosa et al. (2013), “por convención de forma arbitraria por los expertos en el área de estudio” (p. 9). En consecuencia, esta escala no posee un cero absoluto y además el cero no implica la ausencia del atributo. Para las escalas de intervalo se permite el cálculo de proporciones, porcentajes, razones y estadísticos como la media aritmética, mediana, moda, rangos y desviación estándar.

Las escalas de razón “tienen las propiedades de las escalas ordinales y de intervalo, pero además, el cero es real, es absoluto, no es arbitrario” (Coronado, J., 2007, p. 118-119), lo que significa que el cero representa ausencia de la característica. Es así, como los números se pueden comparar por medio de razones (cocientes) y tiene sentido indicar, por ejemplo, cuántas veces el valor de x es más grande que el valor de y . Adicionalmente, en esta escala de medida sólo es arbitraria la unidad de medida y se permite el cálculo de todo tipo de operaciones aritméticas válidas.

2.2.2. Tablas Estadísticas

En el análisis descriptivo se generan tablas y gráficos para resumir información de las variables de manera compacta y precisa, y evidenciar el comportamiento de éstas de manera individual y conjunta. Un caso particular de tablas estadísticas son las **tablas de frecuencias**. Se elaboran teniendo en cuenta un conjunto con n individuos asociados a una variable cuyos valores se han agrupado en k categorías o intervalos denotados con C_1, C_2, \dots, C_k , con $k < n$. Para cada una de las anteriores categorías se definen las siguientes magnitudes que se pueden utilizar con variables cuantitativas o sólo algunas de ellas si la variable es cualitativa:

La **frecuencia absoluta** de la clase C_i es la cantidad de observaciones que pertenecen a la i -ésima categoría o intervalo para $i = 1, 2, \dots, k$. Se denota con f_i .

La **frecuencia relativa** h_i de la clase C_i es la proporción de la frecuencia absoluta de la i -ésima clase con respecto al total de observaciones, es decir

$$h_i = \frac{f_i}{n}, \text{ para } i = 1, 2, \dots, k$$

La **frecuencia acumulada** F_i de la clase C_i es la cantidad de individuos cuya modalidad es inferior o igual a la i -ésima clase, es decir:

$$F_i = \sum_{j=1}^i f_j, \text{ con } i = 1, 2, \dots, k.$$

La **frecuencia relativa acumulada** H_i de la clase C_i es la proporción de individuos con respecto al total de observaciones cuya modalidad es inferior o igual a la i -ésima clase, esto es,

$$H_i = \frac{F_i}{n}, \text{ para } i = 1, 2, \dots, k$$

La tabla que contiene estas categorías con sus correspondientes frecuencias se denomina **distribución de frecuencias** y se utiliza para presentar de forma ordenada la distribución de los datos, con un formato semejante al presentado en la tabla 1.

Tabla 1
Formato de una tabla de distribución de frecuencias.

C_i	f_i	h_i	F_i	H_i
C_1	f_1	h_1	F_1	H_1
C_2	f_2	h_2	F_2	H_2
...
C_k	f_k	h_k	F_k	H_k
Total	n	1		

Proposición. En una tabla de distribución de frecuencias de una categoría con k clases se pueden verificar las siguientes propiedades:

$$i) \sum_{i=1}^k h_i = 1 \qquad ii) F_k = n \qquad iii) H_i = \sum_{k=1}^i h_k \qquad iv) H_k = 1$$

Nota: Para la elaboración de tablas de frecuencias con variables cualitativas es necesario tener en cuenta la escala de medición e indicar la pertinencia del uso de las frecuencias acumuladas y las frecuencias relativas acumuladas. Por ejemplo, si la escala de medición de la variable es nominal, entonces las frecuencias acumuladas carecerían de sentido. Por otra parte, si la escala de medición de la variable es ordinal, entonces es pertinente el uso de las frecuencias acumuladas.

2.2.2.1. Tablas de contingencia para dos variables cualitativas

Se debe agregar que cuando se organizan los datos de una investigación para estudiar la relación entre dos variables cualitativas, es pertinente cambiar la tabla de frecuencias por otra forma de resumir datos categóricos como la **tabla de contingencia bidimensional**, donde X y Y son dos variables categóricas con I niveles para X y J niveles para Y , de manera que un objeto puede estar clasificado en una de las $I \times J$ categorías, como se muestra en la tabla 2 (Aguilera, 2005, p. 3).

Tabla 2
Formato general de una tabla de contingencia $I \times J$

		Factor Y					$n_{i.}$	
		Y_1	...	Y_j	...	Y_J		
Factor X	X_1	n_{11}	...	n_{1j}	...	n_{1J}	$n_{1.}$	
	
	X_i	n_{i1}	...	n_{ij}	...	n_{iJ}	$n_{i.}$	
	
	X_I	n_{I1}	...	n_{Ij}	...	n_{IJ}	$n_{I.}$	
		$n_{.j}$	$n_{.1}$...	$n_{.j}$...	$n_{.J}$	n

Donde:

$i = 1, 2, \dots, I, j = 1, 2, \dots, J$ y n_{ij} representa el número de individuos de la muestra que se clasifican simultáneamente en el nivel X_i de X y Y_j de Y .

$$\text{Factor X: } n_{i.} = \sum_{j=1}^J n_{ij}$$

$$\text{Factor Y: } n_{.j} = \sum_{i=1}^I n_{ij}$$

Nota: Los temas de relación entre dos variables cuantitativas no están al alcance de este nivel de secundaria (correlación y regresión lineal simple).

2.2.2.2. Tablas de frecuencias para una variable cuantitativa

A diferencia de las variables cualitativas nominales, en el estudio de variables cuantitativas (por medio del análisis univariado para este nivel de secundaria) se resume la información utilizando las tablas de distribución de frecuencias (ver tabla 1). Sin embargo, “cuando una variable es cuantitativa de razón, no se dispone directamente de un conjunto de k clases para elaborar la tabla de distribución de frecuencias” (Sosa et al., 2013, p. 22). Y aunque no hay universalidad para determinar el número k de intervalos (clases) de agrupación, en este trabajo se sugiere utilizar la siguiente **regla de Sturges** (teniendo en cuenta que esta regla es utilizada en el software estadístico R):

$$k = [1 + 3.32 \log_{10}(n)]$$

donde n es el tamaño de la muestra y $[\]$ es la función techo definida como el menor entero mayor o igual que el número dado; ejemplos:

$$\text{Ejemplo 1. } [4.7] = 5$$

$$\text{Ejemplo 2. } [3.1] = 4$$

Nota: Ningún intervalo debe tener frecuencia absoluta cero. En caso de presentarse esta novedad, se recurre a $k - 1$ intervalos, o menos intervalos hasta tener frecuencias absolutas de uno o más (Grajales, 2016).

Para determinar la amplitud de los intervalos, es necesario conocer el **rango** R_X de la variable X que se calcula utilizando la siguiente expresión:

$$R_X = x_{m\acute{a}x} - x_{m\acute{i}n}$$

donde $x_{m\acute{a}x}$ representa el valor **máximo** de X y $x_{m\acute{i}n}$ representa el valor **mínimo** de X .

Luego, la expresión para calcular la **amplitud** a es:

$$a = \frac{R_X}{k}, \text{ donde } k \text{ es el número de intervalos.}$$

Los intervalos (o las clases) de agrupación se establecen de la siguiente forma:

$$C_1 = \{x: l_0 \leq x < l_1\} \text{ donde } l_0 = x_{m\acute{i}n} \text{ y } l_1 = l_0 + a$$

$$C_2 = \{x: l_1 \leq x < l_2\} \text{ donde } l_2 = l_1 + a$$

...

$$C_i = \{x: l_{i-1} \leq x < l_i\} \text{ donde } l_i = l_{i-1} + a$$

...

$$C_k = \{x: l_{k-1} \leq x \leq x_{m\acute{a}x}\}.$$

2.2.3. Gráficos Estadísticos

Teniendo en cuenta las características de estudio y el tipo de variable estadística, se pueden utilizar diferentes gráficos estadísticos como diagrama de barras, diagrama de sectores, o histogramas. El objetivo principal de estos gráficos es “dar a entender de manera clara y sencilla el comportamiento de una o varias variables e identificar fácilmente cualquier fenómeno de interés” (Sosa et al., 2013, p. 31).

2.2.3.1. Gráficos para variables cualitativas

Cuando la variable es cualitativa, es común utilizar diagrama de barras y diagrama de sectores.

Diagrama de Barras. En este gráfico se utilizan dos ejes perpendiculares de modo que en uno de ellos se ubican las distintas modalidades de la variable (cada categoría se representa mediante una barra) y en el otro se representan los valores de la frecuencia absoluta o relativa por medio de una recta numérica con valores positivos, comenzando desde el cero para no exagerar diferencias entre las categorías. En el eje donde están ubicadas cada una de las categorías, se levantan las barras con una altura proporcional a su respectiva frecuencia absoluta o relativa, resaltando que las barras difieren sólo en altura y no en ancho (ver sección 2.2.2). Para diferenciar este diagrama de datos categóricos con los de una variable cuantitativa, se sugiere utilizar un diagrama de barras horizontal ubicando las distintas categorías en el eje vertical y las respectivas frecuencias en el eje horizontal (Grajales, 2016).

Cabe resaltar que este tipo de gráfico también se puede utilizar para variables cuantitativas discretas, siempre y cuando el número de categorías sea pequeño.

El **diagrama de Pareto** es un gráfico de barras que se utiliza exclusivamente para variables cualitativas de escala nominal, cuyo objetivo es identificar él o los aspectos prioritarios que hay que tratar en un estudio (con respecto a las modalidades de la variable), así como sus causas más importantes. Para la construcción del diagrama se elabora una tabla con la frecuencia de cada categoría, su porcentaje, frecuencia acumulada y porcentaje acumulado. Luego se construye un gráfico de barras, ordenando las categorías de mayor a menor frecuencia, de izquierda a derecha, para identificar la categoría que predomina o aquellos factores (modalidades de la variable) que tienen mayor importancia.

Diagrama de sectores. Para este gráfico, se divide un círculo en k regiones (con vértice en el centro del círculo), donde k es el número de categorías C_1, C_2, \dots, C_k para una muestra de tamaño n , de manera que a cada categoría le corresponderá un sector del círculo proporcional a su frecuencia absoluta (f_i) o relativa (h_i). Nótese que, el sector del círculo y la frecuencia absoluta o relativa, son magnitudes directamente proporcionales, por lo que se puede establecer las siguientes relaciones:

$$\frac{n}{f_i} = \frac{360^\circ}{v_i}, \quad \frac{1}{h_i} = \frac{360^\circ}{v_i}$$

donde v_i representa el ángulo asociado con el sector circular de la i –ésima categoría para $i = 1, \dots, k$. Así, el arco de cada región se puede calcular utilizando una de las siguientes expresiones, ya sea para representar la frecuencia absoluta o la frecuencia relativa, respectivamente:

$$v_i = \frac{f_i \times 360^\circ}{n}, \quad v_i = h_i \times 360^\circ$$

2.2.3.2. Gráficos para variables cuantitativas

Para variables cuantitativas, se puede utilizar un **diagrama diferencial** donde se representa las frecuencias absolutas o relativas, ó un **diagrama integral** donde se representa las frecuencias absolutas o relativas acumuladas.

Gráficos para variables cuantitativas discretas. Para este tipo de variable se utiliza el diagrama de barras por medio de un diagrama diferencial donde las barras son estrechas para indicar que se trata de un valor discreto (siempre y cuando el número de categorías sea pequeño).

Gráficos para variables cuantitativas continuas.

Histograma. El histograma se construye teniendo en cuenta la distribución de frecuencias (ver sección 2.2.2.2). Se utilizan barras adosadas verticales, de manera que a cada clase le corresponde un rectángulo que tiene como base a su respectivo intervalo (las bases se ubican en el eje horizontal que contiene los valores del rango de la variable). La altura de cada rectángulo es proporcional a la frecuencia de cada intervalo de clase.

Polígono de frecuencias. Para su construcción se tienen en cuenta los rectángulos elaborados en el histograma, para luego unir mediante líneas rectas los puntos superiores de cada rectángulo localizados en los puntos medios de cada intervalo. Es decir, se tienen k puntos (x_i, f_i) y se unen con líneas rectas (con $i = 1, 2, \dots, k$), donde f_i representa la

frecuencia absoluta y x_i representa los puntos medios que se denominan **marcas de clase**. Para calcular las marcas de clase se utiliza la siguiente expresión:

$$x_i = \frac{l_{i-1} + l_i}{2}$$

donde l_{i-1} representa el **límite inferior** y l_i el **límite superior** de la clase o intervalo, para $i = 1, 2, \dots, k$.

Nota: En ocasiones, la gráfica de un polígono de frecuencias se elabora en conjunto con un histograma (para un mismo conjunto de datos) teniendo en cuenta que utilizan la misma escala de medición.

Un **polígono de frecuencias acumulado** u **ojiva** es un diagrama integral que se obtiene de la misma forma que un polígono de frecuencias corriente, a diferencia que se construye utilizando las frecuencias acumuladas en lugar de las frecuencias absolutas.

El **diagrama Boxplot (Caja y bigotes)** se utiliza para representar medidas de localización, dispersión, la simetría y la detección de datos atípicos. Para la construcción de este gráfico es necesario calcular algunos percentiles.

El p – ésimo **percentil** (p_p) es un valor tal que al menos el $p\%$ de los datos es menor o igual que p_p y al menos el $(100 - p)\%$ de los datos es mayor o igual que p_p (Sosa et al., 2013, p. 53).

Un procedimiento para calcular el p – ésimo percentil de un conjunto de n datos no agrupados, según Sosa et al. (2013, p. 53), es el siguiente:

1. Ordenar los datos ascendentemente.
2. Calcular el índice i a través de la fórmula

$$i = \frac{np}{100}$$

donde p es el percentil de interés y n el tamaño de la muestra.

3. Calcular el percentil de acuerdo con uno de los siguientes casos
 - Si el índice i no es entero, se redondea al siguiente entero.
 - Si i es entero, el p – ésimo percentil es el promedio de los valores de los datos ubicados en las posiciones i e $i + 1$.

Si los datos están agrupados en una tabla de frecuencias por intervalos, se realiza los siguientes pasos:

1. Identificar el intervalo $[l_{i-1}, l_i]$ que contiene el p – ésimo percentil, utilizando la siguiente relación

$$i = \text{mín}\{j: F_j > (p\%)n\}.$$

2. Utilizar la siguiente expresión para calcular el p – ésimo percentil:

$$p_p = l_{i-1} + \frac{(p\%)n - F_{i-1}}{f_i} (l_i - l_{i-1})$$

donde F_i es la frecuencia acumulada de la clase i y f_i su frecuencia absoluta.

Dentro de los percentiles existen algunos casos particulares denominados **cuartiles** que dividen el conjunto de datos en cuatro partes con el mismo porcentaje de datos. Los cuartiles q_1 , q_2 y q_3 representan a los percentiles p_{25} , p_{50} y p_{75} , respectivamente. Nótese que el cuartil q_2 y el percentil p_{50} representan a la vez a la mediana, una de las medidas de tendencia central.

A partir de los cuartiles, se puede calcular el **rango intercuartílico** RI_x de x_1, x_2, \dots, x_n -un conjunto de n realizaciones de una variable X -, como la diferencia entre el tercer y el primer cuartil, es decir $RI_x = q_3 - q_1$. Adicionalmente, se pueden calcular el **límite inferior (LI)** y el **límite superior (LS)** de los datos con las siguientes expresiones:

$$LI = q_1 - (1.5)RI_x; \quad LS = q_3 + (1.5)RI_x$$

Los datos menores que LI se conocen como datos **atípicos pequeños** y los datos mayores que LS se conocen como datos **atípicos grandes** (Grajales, 2016).

Teniendo en cuenta los anteriores conceptos, en el boxplot se representan los tres cuartiles y los datos atípicos con base en un rectángulo alineado vertical u horizontalmente. Para su construcción, según Sosa et al. (2013, p. 64), se utilizan los siguientes pasos:

- i. Calcular los cuartiles q_1 , q_2 y q_3 de un conjunto de datos.
- ii. Dibujar un rectángulo paralelo al eje y y tal que su lado mayor inicie en q_1 y termine en q_3 [Aunque el ancho de la caja no es de relativa importancia, Sosa et al. recomienda que sea menor que el largo del rectángulo. Sin embargo, el software estadístico R no cumple necesariamente la anterior condición, de manera que, por cuestiones educativas, en esta estrategia didáctica no se tiene en cuenta tal recomendación].
- iii. Trazar los segmentos de la caja [bigotes], de manera centrada respecto al ancho de la caja, hasta los límites bigote. inf = $\text{máx}\{x_{\text{mín}}, LI\}$ y bigote. sup = $\text{mín}\{x_{\text{máx}}, LS\}$. Por

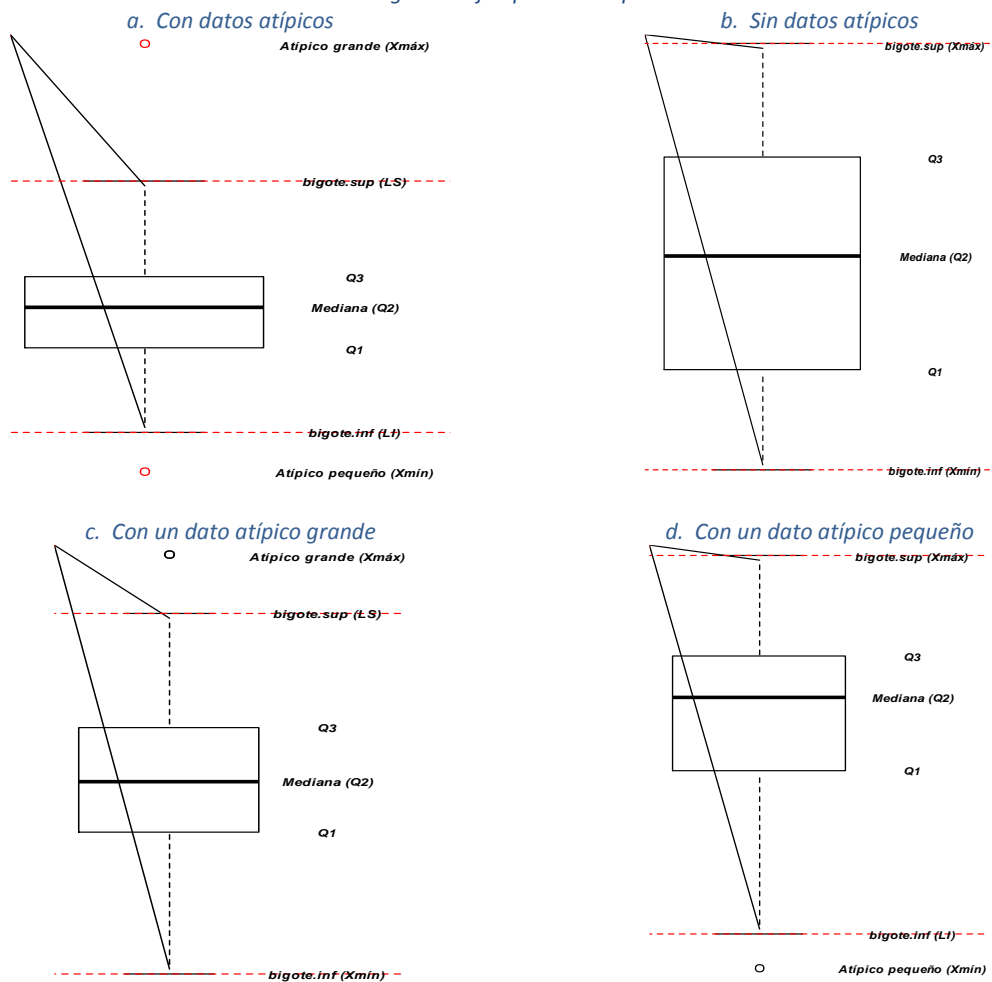
lo tanto, estos segmentos se extienden del bigote inferior (bigote.inf) a q_1 y de q_3 al bigote superior (bigote.sup).

- iv. Señalar y resaltar con un círculo (°) los valores que se encuentran por fuera del intervalo $[LI, LS]$, denominados **datos atípicos**.

Un grupo de datos puede tener ningún, uno o varios datos atípicos, grandes y/o pequeños.

- v. Trazar una línea horizontal dentro de la caja a la altura de q_2 (ver figura 1).

Figura 1: Ejemplos de boxplots.



Finalmente, se debe mencionar que existen otros gráficos para variables cuantitativas continuas como el pictograma, el cartograma, el diagrama de tallos y hojas y el diagrama de líneas, pero, teniendo en cuenta este nivel de secundaria (a quien está dirigido la estrategia didáctica), no se incluyen en esta sección.

2.3. Marco Didáctico: Trabajo con Proyectos

El desarrollo de este marco didáctico se realiza con base en el texto *Estadística con proyectos* (Batanero y Díaz, 2011).

Una de las metodologías de enseñanza en la estadística es el trabajo con proyectos que busca aplicar el conocimiento estadístico en la resolución de problemas de diversas disciplinas, considerando que la estadística está ligada a las aplicaciones y que el razonamiento estadístico es una herramienta de la resolución de problemas. Durante el trabajo con proyectos en el aula se busca contextualizar la estadística, hacerla relevante y despertar el interés en el estudiante hacia un problema de investigación, contribuyendo a la adquisición de competencias en comunicación, en matemáticas, en la interacción con el mundo físico, en el tratamiento de la información, entre otras. Más específicamente, se busca mejorar la capacidad de argumentación y formulación de conjeturas, y hacer uso del análisis exploratorio de datos por medio de cambios de registros de representación de un gráfico a otro.

Para la realización de estos proyectos se busca cuestionar al estudiante acerca del problema que desea indagar (un problema abierto, realista y apropiado a su nivel, que despierte su interés), los datos que necesita, la forma de recolectar la información y el significado que tiene el resultado en su práctica. Por tanto, la utilización del trabajo con proyectos es catalogado como una auténtica investigación que afronta el estudiante a través de las distintas fases de “planteamiento de un problema, decisión sobre los datos a recoger, recogida y análisis de datos y obtención de conclusiones sobre el problema planteado” (Batanero y Díaz, 2011, p. 13).

Una de las fases iniciales del trabajo con proyectos es elegir la pregunta de investigación que debe ser manifestada por el estudiante y orientada por el profesor para dar paso al problema, teniendo en cuenta los siguientes aspectos: “¿Qué quieres probar?, ¿Qué tienes que medir?, ¿Qué datos necesitas?, ¿Cómo encontrarás tus datos?, ¿Qué harás con ellos?, ¿Crees que puedes hacerlo?, ¿Podrás contestar tu pregunta? [y] ¿Para qué te servirán los resultados?” (Batanero y Díaz, 2011, p. 23).

Durante la realización del proyecto, es necesario animar a los estudiantes a escribir su informe acerca de su “investigación” de forma clara y lógica, porque además de desarrollar la habilidad de producción de informes comprensivos y estructurados que incorporen el uso de información estadística como forma de argumentación, se refuerza el proceso de

razonamiento estadístico de relatar a otra persona sus decisiones, acciones e interpretaciones.

A su vez, para el proceso de evaluación del proyecto (cuyo principal propósito es mejorar el aprendizaje de los estudiantes) se deben tener en cuenta los distintos aspectos del conocimiento matemático como comprensión conceptual (pasar de un modo de representación a otro), conocimiento procedimental (reconocer procedimientos correctos o incorrectos), resolución de problemas, formulación y comunicación matemática, razonamiento matemático (identificar patrones y conjeturas) y actitud hacia las matemáticas.

Pero además de evaluar aspectos del conocimiento matemático, se puede tener en cuenta la relevancia de la pregunta de interés, el diseño de la investigación con énfasis en la población y la muestra, el análisis de datos con respecto a la pregunta de investigación, las conclusiones referentes al análisis, la reflexión del proceso realizado, la presentación de los resultados con énfasis en la claridad de los gráficos y la creatividad, y la originalidad del estudiante.

2.4. Marco Metodológico: Investigación Acción

Este trabajo se desarrolla con base en la investigación acción como metodología orientada hacia el cambio educativo que se construye desde y para la práctica en la búsqueda de mejorar la labor docente a partir del análisis crítico de las situaciones y se configura en los ciclos de planificación, acción, observación y reflexión (Bausela, 2004).

El carácter de esta investigación es mixta porque integra los enfoques cualitativo y cuantitativo (Gómez, 2006). Por un lado, en el enfoque cualitativo se realiza un seguimiento a los procesos académicos de los estudiantes respecto a utilización de gráficos estadísticos. Por otro lado, se complementa con aspectos cuantitativos por medio de la comparación entre los conocimientos previos del estudiante y los resultados alcanzados al finalizar el proyecto.

Las variables de estudio en esta investigación están asociadas a la construcción, lectura, interpretación y análisis de gráficos estadísticos por parte de los estudiantes. Específicamente, se estudiará la pertinencia del uso de un gráfico, la escala numérica que utiliza, la proporción de los ejes coordenados, el cálculo de porcentajes, entre otros.

Capítulo 3: Estrategia didáctica

En el diseño de esta estrategia didáctica se elabora una prueba diagnóstica con la finalidad de reconocer los conocimientos previos que tienen los estudiantes respecto al manejo de gráficos estadísticos, para luego estructurar un proyecto con preguntas dirigidas hacia la construcción, interpretación y análisis de estos gráficos aplicados en un contexto real. Luego de desarrollar este proyecto en clase se implementa nuevamente esta prueba con algunas modificaciones. Con ello, podemos identificar el efecto que tuvo la estrategia didáctica con respecto a los conocimientos estadísticos adquiridos por los estudiantes.

3.1. Prueba diagnóstica

Para identificar los conocimientos previos de los estudiantes con respecto a la construcción, lectura, interpretación y análisis de información que aparece en tablas y gráficos estadísticos, se diseña una prueba diagnóstica tomando como referencia modelos de preguntas propuestos en *Los lineamientos para las aplicaciones muestral y censal 2013* (ICFES, 2013), entrevistas con el profesor Luis F. Grajales H. y notas de clase del curso Métodos estadísticos de la Maestría en Enseñanza de las Ciencias Exactas y Naturales, por parte del profesor Ramón Giraldo en el año 2016.

La prueba diagnóstica está conformada por diez preguntas de las cuales las primeras siete son de opción múltiple, es decir, contienen un enunciado y cuatro opciones de respuestas (A, B, C, D). Entre las opciones presentadas existe sólo una que satisface las exigencias del enunciado. Adicionalmente, las últimas tres preguntas son abiertas (ver anexo A. Prueba diagnóstica).

Análisis a priori de la prueba diagnóstica

A cada una de las preguntas de la prueba se le elabora una afirmación, enunciados globales con respecto a los conocimientos de los estudiantes, sus capacidades o habilidades (ICFES, 2013), y se realiza una breve descripción de la respuesta correcta (ver tabla 3).

Tabla 3
Análisis a priori de la prueba diagnóstica

Pregunta 1	Afirmación: Reconocer las relaciones entre los niveles de una variable categórica, en una tabla de frecuencias, y su representación gráfica.
-------------------	---

	<p>Para responder correctamente esta pregunta se deben conocer los convenios de construcción y elementos de un gráfico de barras vertical para los niveles de una variable categórica como: i) Interpretar la información de la tabla de frecuencias, ii) construir gráficos de barras con una escala proporcional lineal con origen en cero y; iii) asociar la altura de cada barra con su respectiva frecuencia absoluta.</p>
Pregunta 2	<p>Afirmación: Resolver problemas a partir de un conjunto de datos presentados en una tabla de contingencia.</p> <p>Para responder correctamente esta pregunta es necesario interpretar los datos de una tabla de contingencia, reconocer las variables categóricas y, para calcular el total, realizar la sumatoria de los totales de cada categoría o la sumatoria de todas las celdas de la tabla.</p>
Pregunta 3	<p>Afirmación: Comparar e interpretar analíticamente los niveles de una variable categórica en una tabla de contingencia.</p> <p>Para responder correctamente esta pregunta es necesario reconocer los niveles de una variable categórica, calcular los totales de cada categoría y comparar sumas apropiadamente (entre algunos de estos totales). En este caso, interpretar los datos de la variable categórica correspondiente a <i>Música preferida</i> y realizar la sumatorias de datos por niveles teniendo en cuenta la clasificación por región.</p>
Pregunta 4	<p>Afirmación: Calcular porcentajes para representar información en contextos de medida.</p> <p>Para responder correctamente esta pregunta es necesario reconocer la relación “<i>parte-todo</i>” entre la frecuencia por género musical y el total de estudiantes. Conocer y dominar el concepto de porcentaje y, además, es necesario realizar operaciones básicas matemáticas para calcular porcentajes.</p>
Pregunta 5	<p>Afirmación: Comparar gráficas de barras verticales adecuadas para presentar diversos tipos de datos.</p> <p>Para responder correctamente esta pregunta es necesario interpretar la información que aparece en la tabla de frecuencias, identificar la columna que se pide graficar y tener en cuenta los elementos para la construcción de un gráfico de barras vertical como los ejes, las escalas y las etiquetas de los ejes.</p>

Pregunta 6	<p>Afirmación: Producir y comparar representaciones gráficas para presentar datos, comprendiendo que distintas representaciones de los mismos datos pueden conducir a distintas interpretaciones.</p>
	<p>Para responder correctamente esta pregunta y realizar un cambio de registro (representación) es necesario conocer: i) la representación de la unidad, el 100%, como la región de área del círculo que corresponde a un ángulo de 360° y las fracciones de esta unidad (fracciones del ángulo), ii) el procedimiento de construcción de un gráfico de barras y de un diagrama circular; iii) realizar una correspondencia entre los datos “leídos” del gráfico de barras con su respectiva representación en el diagrama circular y iv) tener en cuenta los elementos para la construcción de un gráfico como etiquetas, el área de una región circular y las convenciones.</p>
Pregunta 7	<p>Afirmación: Usar porcentajes para resolver problemas en contextos de medida. Comprender y calcular incrementos y reducciones porcentuales en contexto.</p>
	<p>Para responder correctamente esta pregunta el estudiante debe saber calcular el porcentaje del total de uno o más niveles de una variable categórica.</p>
Pregunta 8	<p>Afirmación: Interpretar y producir representaciones gráficas adecuadas para ilustrar diversos tipos de datos presentados en una tabla de contingencia.</p>
	<p>Para responder correctamente esta pregunta el estudiante debe saber leer una tabla de contingencia, analizar los totales de cada nivel de la variable categórica y usar esta información para determinar los datos faltantes por medio de algunos cálculos. Además, debe conocer el procedimiento de construcción de un gráfico circular para representar los datos pedidos. Es fundamental conocer y aplicar el concepto de porcentaje, como también la relación entre el 100% de los datos y los 360° del círculo.</p>
Pregunta 9	<p>Afirmación: Interpretar y producir representaciones gráficas adecuadas para presentar diversos tipos de datos.</p>
	<p>Para responder correctamente esta pregunta el estudiante debe saber leer una tabla de contingencia, extraer información parcial de ella y conocer el procedimiento de construcción de un gráfico de barras.</p>

Pregunta 10	Afirmación: Interpretar y producir representaciones gráficas adecuadas para presentar diversos tipos de datos.
	Para responder correctamente esta pregunta el estudiante debe saber leer una tabla de contingencia, extraer información parcial de ella, conocer el procedimiento de construcción de un gráfico de barras y calcular porcentajes con respecto a los niveles de la variable (categórica) de análisis.

3.2. El proyecto: Estadísticas de sexualidad

Para la elaboración y estructuración de este proyecto se tomaron como referencia algunos modelos propuestos en el texto *Estadística con proyectos* (Batanero y Díaz, 2011).

3.2.1 Objetivos del proyecto

Realizar la lectura, interpretación y análisis de datos estadísticos con respecto a la percepción de educación sexual, comprendiendo la importancia y utilidad de la recolección de la información, y analizar la relación que existe entre diferentes variables.

Además, se busca mostrar la utilidad de la estadística descriptiva en el estudio de situaciones cotidianas por medio de los cambios de registros de representación -utilizando tablas y gráficos estadísticos-, el método de análisis de los datos según el tipo de variable, el estudio de interrelaciones entre variables, entre otros.

3.2.2. Los datos

Para el diseño de la actividad se utiliza la información de la Encuesta Nacional de Demografía y Salud [ENDS], un proyecto desarrollado por Profamilia desde 1990 con actualizaciones quinquenales. En el año 2010, durante un estudio realizado acerca de la salud sexual y reproductiva se entrevistaron a 53521 mujeres en edad fértil (13-49 años). No obstante, para este ejercicio académico con los estudiantes, se utiliza una sub-base con sólo 80 registros, con edades entre los 25 y 30 años, y con residencia en Bogotá.

Es necesario resaltar que, al estudiar las preguntas de esta encuesta se observaron algunos errores técnicos como el trasfondo machista en la base de datos con respecto a la actitud de la mujer de tener un hijo homosexual dado que no se tiene en cuenta la posible

condición de hijas homosexuales. Además, debido a que se aborda el tema de sexualidad con los estudiantes, fue procedente solicitar la asesoría de Adriana Mosquera P., psicóloga y orientadora del colegio, para estudiar la pertinencia de las preguntas a utilizar en las actividades y evitar temas álgidos que puedan herir susceptibilidades.

En la tabla 4 se muestran algunos de los datos presentados a los estudiantes para el desarrollo de la actividad (Ver Anexo B. Hoja de códigos: Estadísticas de sexualidad).

Tabla 4
Datos de planificación familiar y percepción de la educación sexual

Código	Edad	N_Edu	A_Acad	...	M_Ant	Peso	Altura	E_PR	E_TS
1	30	0	0		1	59.0	162.5	16	1
2	25	1	3		11	60.9	153.2	17	1
3	28	1	4		2	49.3	140.8	15	1
...									

3.2.3. La actividad

Inicialmente, se solicita a los estudiantes discutir el significado de las variables estadísticas incluidas en este estudio, su pertinencia y utilidad para la toma de decisiones. Cabe resaltar que, estos estudiantes se encuentran vinculados con el proyecto de educación sexual que organiza la orientación y el área de matemáticas del colegio, de manera que (con antelación) investigan temas relacionados con el ciclo ovulatorio de la mujer, métodos anticonceptivos, enfermedades de transmisión sexual [ETS], diversidad sexual, entre otros.

Con la información que se entrega (los datos impresos, los datos enviados en archivo de Excel a los correos electrónicos de los estudiantes y la hoja de códigos impresa), los estudiantes trabajan durante cuatro sesiones, en grupos de cuatro integrantes, las preguntas del proyecto (cada proyecto contiene cinco preguntas y sólo las tres primeras son comunes para todos los grupos) para, luego, entregar un informe con la libertad de escoger las tablas y los gráficos estadísticos que consideren convenientes y con la posibilidad de utilizar un software estadístico (Ver Anexo C. Proyecto. Estadísticas de sexualidad). A continuación, se presenta un análisis a priori para cada una de las preguntas utilizadas en los proyectos teniendo en cuenta el tema que se aborda (referente a estadística descriptiva), su objetivo y una propuesta de solución.

1. ¿Cuál es el método anticonceptivo típico utilizado por las mujeres? ¿Cómo ha llegado a esa respuesta?

Tema: La moda.

Objetivo: Identificar la moda como valor típico para una variable cualitativa nominal.

Propuesta de solución

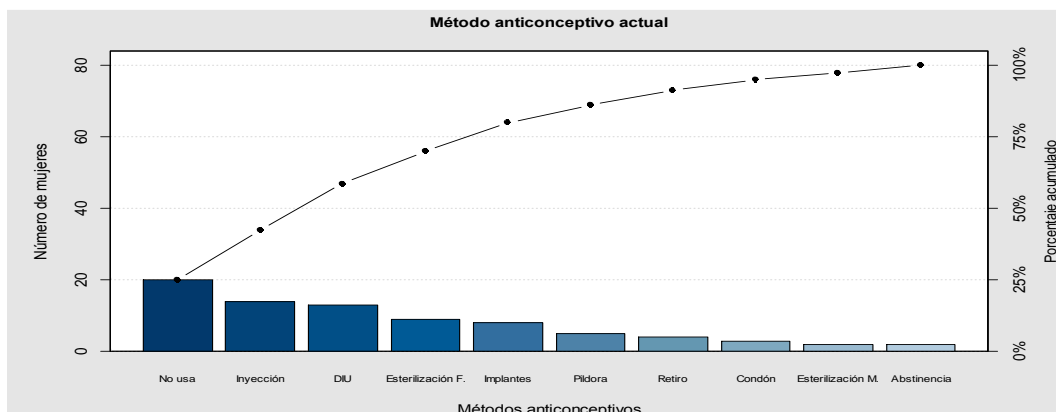
Teniendo en cuenta que el “Método anticonceptivo actual” es una variable cualitativa nominal, en este trabajo se propone realizar una tabla que contenga sus frecuencias absolutas (ver tabla 5), para identificar la categoría con mayor frecuencia (la moda).

Tabla 5: Número de mujeres por cada método anticonceptivo

Código	Método anticonceptivo actual	Número de mujeres
0	No usa	20
1	Píldora	5
2	DIU	13
3	Inyección	14
5	Condón	3
6	Esterilización femenina	9
7	Esterilización masculina	2
8	Periodo de abstinencia	2
9	Retiro	4
11	Implantes	8

Adicionalmente, se puede identificar la moda con la realización de un gráfico estadístico, como el diagrama de Pareto, ordenando las categorías de la variable de mayor a menor frecuencia, de izquierda a derecha, encontrando la categoría que predomina, es decir, la barra con mayor altura que en este caso corresponde a “no usa” y que representa el 25% de los datos (ver figura 2). Se resalta que para el grupo observado lo más común es no usar anticonceptivos. No obstante, como se pregunta por el “método anticonceptivo” y la categoría “no usa” no corresponde a uno de éstos, se elige la “inyección” como la moda.

Figura 2: Gráfico de Pareto para uso de métodos anticonceptivos



2. *¿Existe alguna relación entre el sexo biológico de la cabeza de hogar con lo que haría si se descubre que uno de sus hijos es homosexual? ¿Podría representar la anterior información?*

Nota del docente: Para este trabajo la palabra hijos, en dicha base de datos, hace referencia tanto a hombres como a mujeres.

Tema: Gráfico de barras para una distribución conjunta de dos variables categóricas.

Objetivos: Realizar una tabla de contingencia para dos variables categóricas utilizando porcentajes para las frecuencias absolutas conjuntas, con respecto a su distribución marginal, y representar la información en un gráfico de barras. Nótese que estos porcentajes -entre las dos variables- son comparables entre sí, mientras que las frecuencias absolutas conjuntas no lo son porque pertenecen a diferentes totales (frecuencias marginales).

Propuesta de solución

Se estudia la relación entre las variables “sexo biológico de la cabeza de hogar” y lo que “haría si se descubre que uno de sus hijos es homosexual” (Ver tabla 6) por medio de una tabla de contingencia, calculando la distribución de la reacción si se descubre que uno de sus hijos es homosexual según el sexo biológico de la cabeza de hogar.

Tabla 6: Convenciones usadas respecto a lo que “haría si se descubre que uno de sus hijos es homosexual”

Código	Reacción si un hijo es homosexual
1	Lo retiraría de la casa
3	Lo envía a ver un psicólogo
5	Apoyaría
6	Otro
7	Aceptaría
8	Proporcionaría consejo

Cabe resaltar que las categorías para la variable “sexo biológico de la cabeza de hogar” son hombre o mujer, y que la encuesta fue respondida sólo por mujeres, entonces se utilizará los convenios *jefatura masculina* y *jefatura femenina* para representar *hombre* y *mujer*, respectivamente.

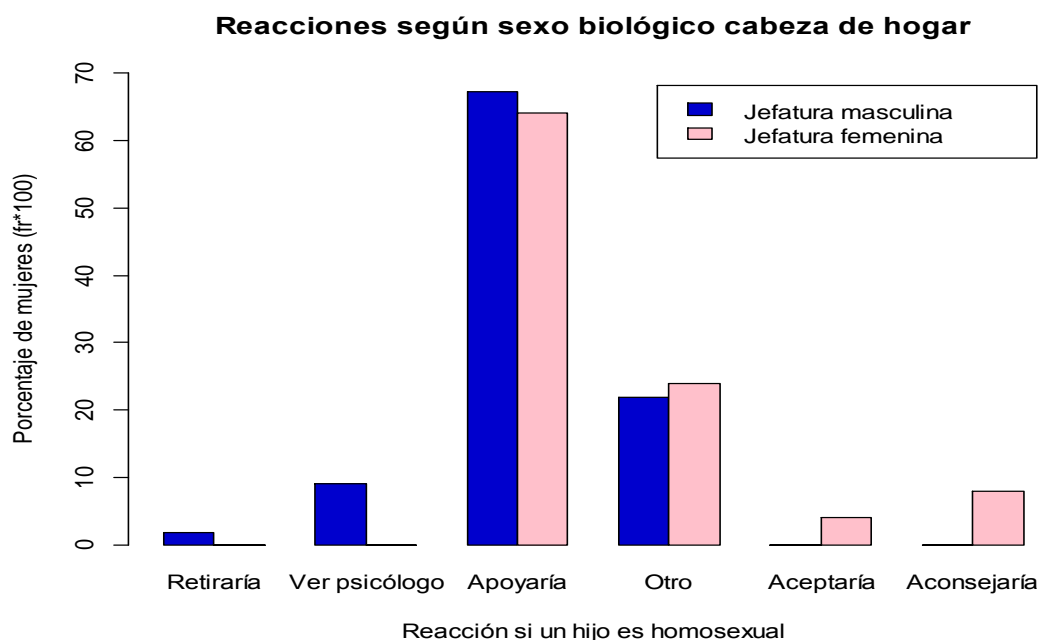
Tabla 7: Clasificación de mujeres según el sexo biológico de la cabeza de hogar y la reacción si se descubre que un hijo es homosexual

	Reacción si un hijo es homosexual						Total
	1	3	5	6	7	8	
Jefatura masculina	1	5	37	12	0	0	55
	1.82%	9.1%	67.26%	21.82%	0%	0%	100%
Jefatura femenina	0	0	16	6	1	2	25
	0%	0%	64%	24%	4%	8%	100%
Total	1	5	53	18	1	2	80

Se puede observar en la tabla 7 que las reacciones de “apoyarlo” (5) o de “otra” alternativa (6) tienen proporciones semejantes tanto para la jefatura masculina como la femenina. Para algunas reacciones más drásticas como “retirarlo de la casa” (1) o “enviarlo a ver un psicólogo” (3), sólo se presentan en hogares con jefatura masculina. De otro lado, las reacciones como “dialogar” (7) o “aconsejar” (8) sólo se presentan en hogares con jefatura femenina.

Las anteriores observaciones se pueden visualizar al representar la información de la tabla 7, utilizando los porcentajes de las frecuencias conjuntas con respecto a las distribuciones marginales (frecuencias relativas condicionadas multiplicadas por 100), por medio de la figura 3.

Figura 3: Reacción si se descubre que un hijo es homosexual según el sexo biológico de la cabeza de hogar



3. ¿Cómo representar la altura de las mujeres entrevistadas?

Tema: Gráfico estadístico para una variable cuantitativa continua.

Objetivo: Realizar un gráfico estadístico apropiado para representar la altura de las mujeres.

Propuesta de solución

La altura de las mujeres es una variable de naturaleza cuantitativa continua. Dado que es la primera exploración de esta variable, se propone realizar un histograma para representar sus datos y observar su comportamiento.

Para la construcción del histograma, se tienen en cuenta los siguientes cálculos:

- El rango de la altura es $R_{\text{Altura}} = x_{\text{máx}} - x_{\text{mín}} = 170.6 - 140.8 = 29.8$
- El número de intervalos sugerido por la regla de Sturges es $k = [1 + 3.32 \log_{10}(80)] = 8$
- La amplitud de cada intervalo es $a = \frac{R_{\text{Altura}}}{k} = \frac{29.8}{8} = 3.725$. Para facilidad de cálculos, se eligen intervalos de amplitud 3.8.

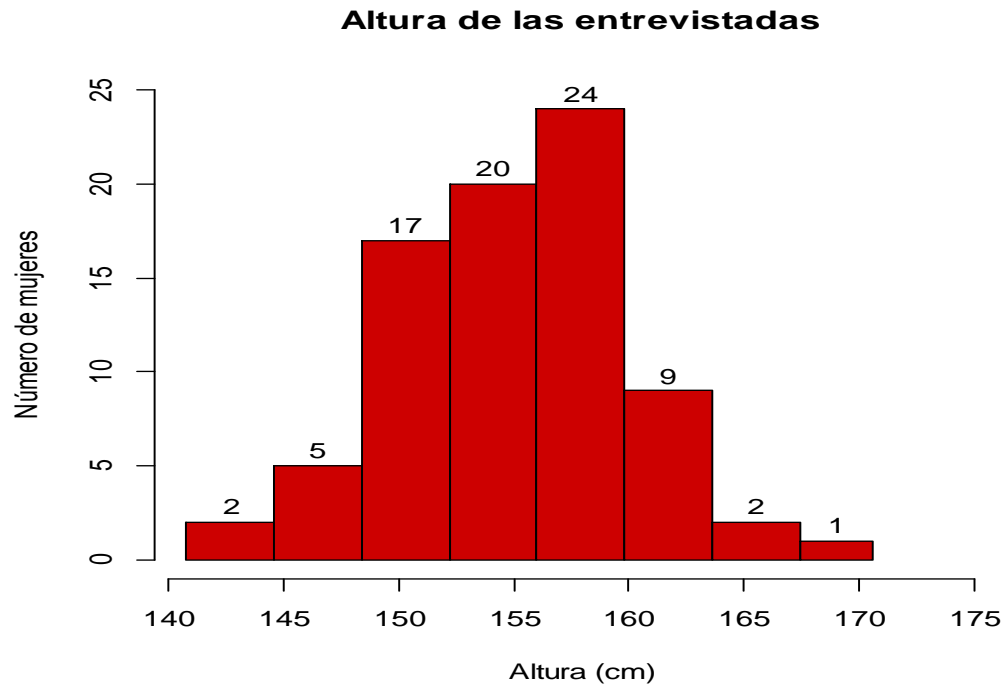
A continuación, se muestra la tabla 8 con la distribución de frecuencias para las alturas:

Tabla 8: Distribución de frecuencias para las alturas

<i>Alturas</i>	f_i	h_i	F_i	H_i
[140.8,144.6)	2	0.025	2	0.025
[144.6,148.4)	5	0.062	7	0.087
[148.4,152.2)	17	0.212	24	0.300
[152.2,156.0)	20	0.250	44	0.550
[156.0,159.8)	24	0.300	68	0.850
[159.8,163.6)	9	0.112	77	0.962
[163.6,167.4)	2	0.025	79	0.987
[167.4,170.6]	1	0.012	80	1.000

Finalmente, se elabora un histograma que representa la altura de las mujeres entrevistadas como se muestra en la figura 4.

Figura 4: Histograma para las alturas



Nota: La solución de esta pregunta -utilizando el software R-Project- se describe en el Apéndice.

4. ¿Existen valores atípicos en la altura de las mujeres entrevistadas? Si los hay, ¿son valores atípicos pequeños o grandes?

Tema: Construcción e interpretación de un boxplot.

Objetivo: Construir un boxplot para identificar posibles valores atípicos con respecto a la altura de las mujeres entrevistadas.

Propuesta de solución

Los valores atípicos se pueden reconocer con facilidad luego de construir un boxplot. De manera que para su construcción se tienen en cuenta los siguientes aspectos:

- Primer cuartil: se identifica el intervalo que contiene el percentil 25, en este caso corresponde a [148.4,152.2). Luego, se reemplazan los valores en:

$$l_{i-1} + \frac{(p\%)n - F_{i-1}}{f_i} (l_i - l_{i-1}). \text{ Se obtiene } q_1 = 148.4 + \frac{20-7}{17} (3.8) \approx 151.3.$$

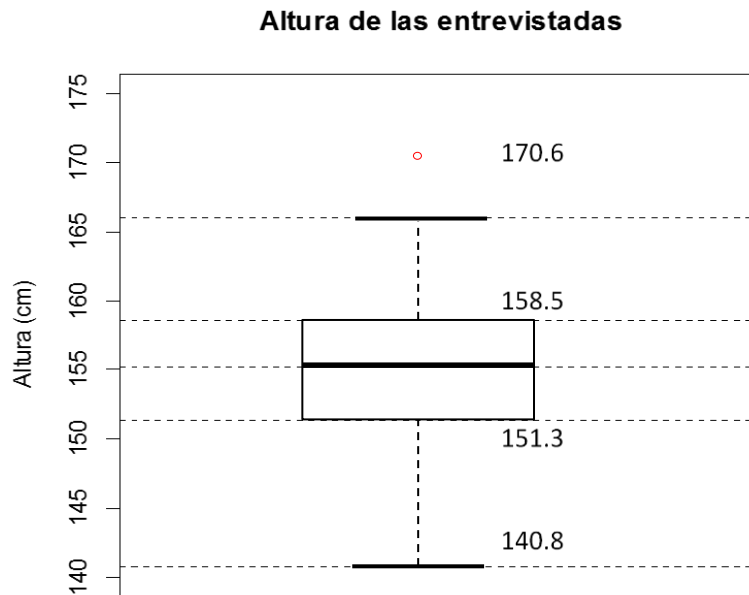
De manera análoga, el segundo y tercer cuartil son $q_2 \approx 155.2$ y $q_3 \approx 158.5$, respectivamente.

Nota: Dada la naturaleza cuantitativa continua de la variable y para ser consecuente con la tabla de distribución de frecuencias del punto anterior, se decide utilizar (en esta propuesta de solución) cálculos para datos agrupados.

- El rango intercuartílico corresponde a: $RI_{\text{Altura}} = q_3 - q_1 \approx 158.5 - 151.3 \approx 7.2$
- El límite inferior del gráfico corresponde a $LI = q_1 - (1.5)RI_{\text{Altura}} \approx 140.5$ y el bigote inferior de la caja a $\text{bigote.inf} = \max\{x_{\min}, q_1 - (1.5)RI_x\} = 140.8$.
- El límite superior del gráfico corresponde a $LS = q_3 + (1.5)RI_{\text{Altura}} \approx 169.3$ y el bigote superior de la caja a $\text{bigote.sup} = \min\{x_{\max}, q_3 + (1.5)RI_x\} = 166$.

Finalmente, se elabora el boxplot para las alturas de las mujeres entrevistadas (ver figura 5).

Figura 5: Boxplot para las alturas



Con el gráfico se puede identificar el valor 170.6 mayor a 169.3 (el límite superior) y, por lo tanto, se considera como un dato atípico grande.

5. ¿Cuál es el intervalo del peso más común entre las mujeres entrevistadas?

Tema: Interpretación de un histograma (para la variable peso, en kg).

Objetivo: Identificar el intervalo modal en un histograma como el rectángulo de mayor altura y determinar la moda como la marca de clase en este intervalo.

Propuesta de solución

De manera análoga al punto tres, se elabora una tabla de distribución de frecuencias para los pesos de las mujeres (ver tabla 9) teniendo en cuenta los siguientes aspectos:

- El rango del peso es $R_{\text{peso}} = 82.5 - 36.0 = 46.5$

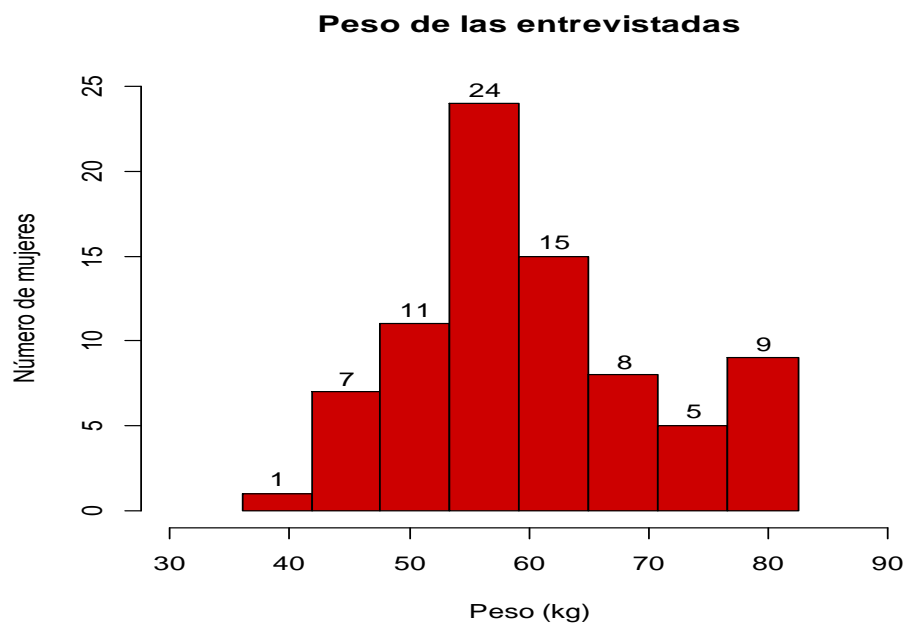
- Se propone utilizar 8 intervalos (teniendo en cuenta la regla de Sturges), la amplitud corresponde a $a = \frac{R_{\text{peso}}}{k} = \frac{46.5}{8} = 5.8125$. Para facilidad de cálculos, se eligen intervalos con amplitud $a = 5.8$.

Tabla 9: Distribución de frecuencias para los pesos

Pesos	f_i	h_i	F_i	H_i	x_i
[36.0,41.8)	1	0.012	1	0.012	38.9
[41.8,47.6)	7	0.087	8	0.100	44.7
[47.6,53.4)	11	0.137	19	0.237	50.5
[53.4,59.2)	24	0.300	43	0.537	56.3
[59.2,65.0)	15	0.187	58	0.725	62.1
[65.0,70.8)	8	0.100	66	0.825	67.9
[70.8,76.6)	5	0.062	71	0.887	73.7
[76.6,82.5]	9	0.112	80	1.000	79.5

Con la tabla 9, se puede identificar el intervalo modal [53.4,59.2), teniendo en cuenta que registra la mayor frecuencia absoluta. Entonces la moda, utilizando el cálculo para la marca de clase del intervalo modal, es $x_i = \frac{l_{i-1} + l_i}{2} = \frac{53.4 + 59.2}{2} = 56.3$. Sin embargo, este intervalo modal puede “hacerse más visible” cuando se construye un histograma (ver figura 6).

Figura 6: Histograma para los pesos



6. ¿Podrías representar en un gráfico la información referente al conocimiento del periodo fértil en el ciclo ovulatorio utilizando términos porcentuales? Y si ahora restringimos la información entre las mujeres con nivel educativo secundaria, ¿podrías realizar el gráfico?

Tema: Gráfico estadístico para una variable cualitativa nominal.

Objetivo: Determinar las cantidades porcentuales de las frecuencias absolutas y representarlas en un gráfico de barras o de sectores.

Propuesta de solución

Teniendo en cuenta que “el conocimiento del periodo fértil en el ciclo ovulatorio” es una variable cualitativa nominal, se organiza la información en una tabla de frecuencias que contenga las cantidades porcentuales (ver tabla 10).

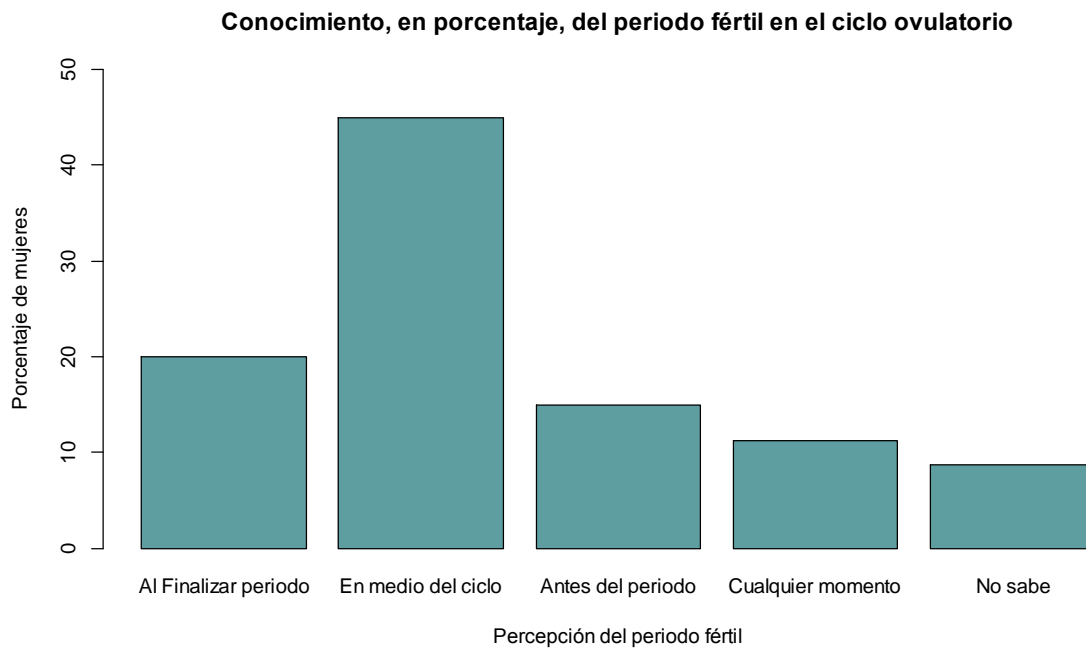
Nota: En la ENDS 2010 “se preguntó a las mujeres en qué momento del ciclo menstrual creían que existe mayor riesgo de quedar embarazadas” teniendo en cuenta que existen parejas que usan el método del ritmo para planificar (Ojeda, Ordoñez y Ochoa, 2011).

Tabla 10: Frecuencias y porcentajes para el conocimiento del periodo fértil en el ciclo ovulatorio

Percepción del periodo fértil	f_i	h_i	Porcentaje
Después de finalizar la menstruación	16	0.20	20
En medio del ciclo	36	0.45	45
Antes de que comience la menstruación	12	0.15	15
En cualquier momento	9	0.11	11
No sabe	7	0.09	9
Total	80	1.00	100

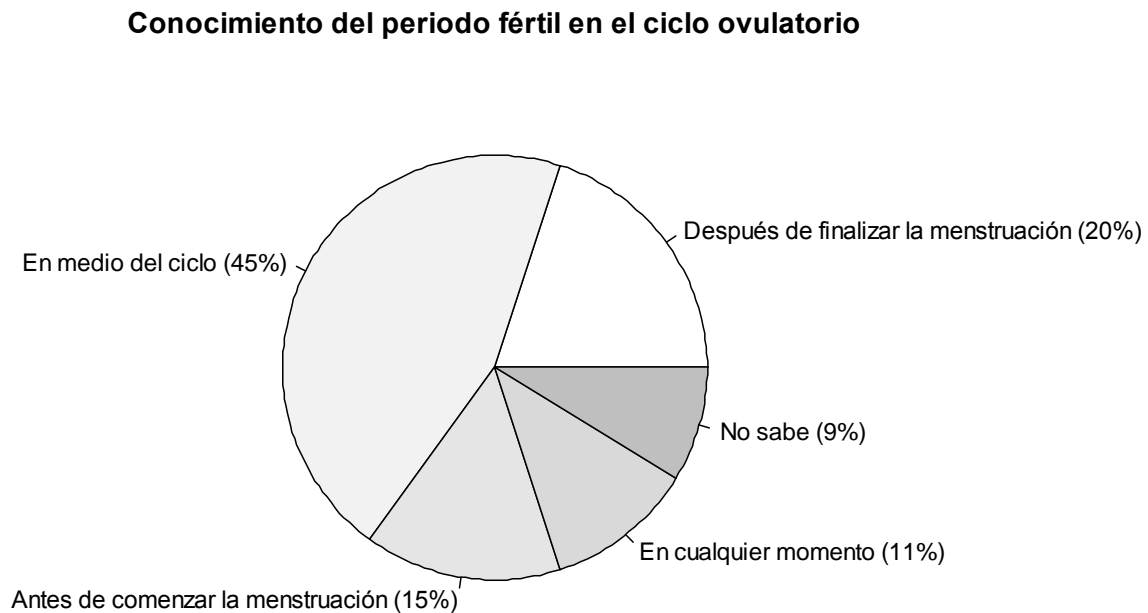
Luego, se elabora un gráfico de barras con los porcentajes de la tabla 10 tal como se muestra en la figura 7.

Figura 7: Diagrama de barras, en porcentaje, para el conocimiento del periodo fértil en el ciclo ovulatorio



Cabe resaltar que también se puede construir un gráfico de sectores para representar la información, tal como se muestra en la figura 8.

Figura 8: Diagrama de sectores para el conocimiento del periodo fértil en el ciclo ovulatorio



Entretanto, para restringir la información entre aquellas mujeres con nivel educativo secundaria, se recomienda realizar una tabla de contingencia entre estas dos variables: “conocimiento del periodo fértil” y “nivel educativo” (ver tabla 11).

Tabla 11: Clasificación de mujeres según el conocimiento del periodo fértil en el ciclo ovulatorio y el nivel educativo

	Nivel educativo más alto				Total
	Sin educación	Primaria	Secundaria	Mayor	
Después de finalizar la menstruación	0	4	6	6	16
En medio del ciclo	0	2	13	21	36
Antes de comenzar la menstruación	0	3	7	2	12
En cualquier momento	0	2	4	3	9
No lo sabe	1	0	3	3	7
Total	1	11	33	35	80

Luego, se realiza una distribución del “conocimiento del periodo fértil en el ciclo ovulatorio” según el “nivel educativo secundaria”, como se muestra a continuación:

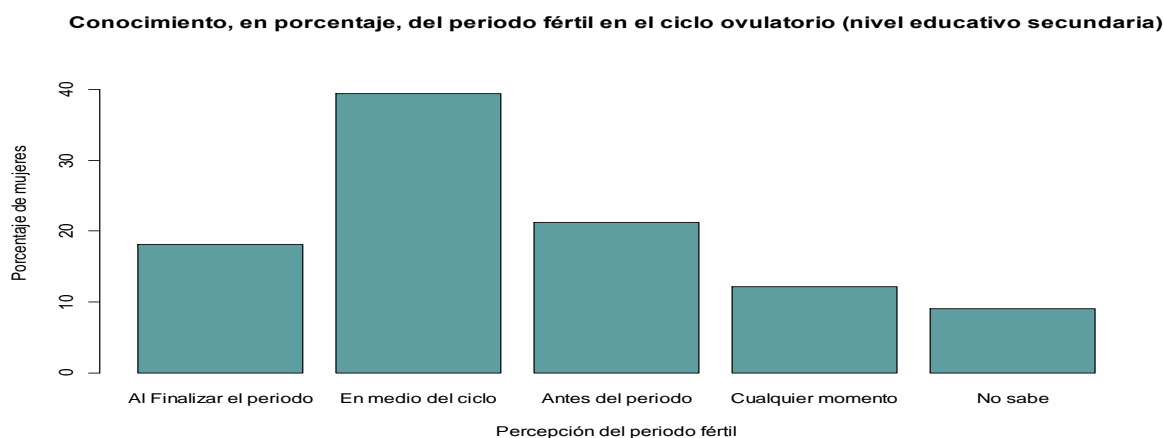
Tabla 12: Frecuencias para el conocimiento del periodo fértil en el ciclo ovulatorio entre las mujeres con nivel educativo secundaria

	Nivel educativo Secundaria
Después de finalizar la menstruación	6 18.18%
En medio del ciclo	13 39.40%
Antes de comenzar la menstruación	7 21.21%
En cualquier momento	4 12.12%
No sabe	3 9.09%
Total	33 100%

Nota: En la tabla 12 se pueden visualizar los porcentajes de cada frecuencia conjunta con respecto a la distribución marginal por columna, utilizando el cálculo de las frecuencias relativas condicionadas (para el nivel educativo secundaria) multiplicadas por 100.

Finalmente, se realiza un gráfico que represente la información del “conocimiento del periodo fértil en el ciclo ovulatorio” entre aquellas mujeres que pertenecen al “nivel educativo secundaria” (ver figura 9).

Figura 9: Diagrama de barras para el conocimiento, en porcentaje, del periodo fértil en el ciclo ovulatorio entre las mujeres que pertenecen al nivel educativo secundaria



7. Calcular las medidas resumen del número de hijos con respecto a cada nivel educativo, usando dos medidas resumen diferentes: la media y la mediana. ¿Qué diferencias observas entre las medidas resumen? ¿Elegirías la media o la mediana para representar la medida resumen del número hijos, según su nivel educativo?

Tema: El efecto de datos atípicos en los cálculos de la media y la mediana.

Objetivo: Identificar una medida resumen pertinente para un conjunto de datos teniendo en cuenta la presencia de datos atípicos.

Propuesta de solución

Se recomienda realizar una tabla de contingencia para relacionar las variables “total de hijos nacidos” y “nivel educativo más alto” (ver tabla 13).

Tabla 13: Clasificación de mujeres según el número de hijos nacidos y el nivel educativo

		Nivel educativo más alto		
		Primaria	Secundaria	Mayor
Total de hijos nacidos	0	0	1	20
	1	2	15	12
	2	6	12	1
	3	2	4	2
	4	0	1	0
	5	1	0	0
Total		11	33	35

Nota: No se incluye el dato de una mujer “sin educación” y con “3 hijos”, pues no permite realizar cálculos y gráficos posteriores como la media y el boxplot.

Luego, se calculan tanto la media como la mediana del número de hijos según el nivel educativo (ver tabla 14).

Tabla 14: Media y mediana del número de hijos nacidos según el nivel educativo

Nivel educativo	Media: Total de hijos	Mediana: Total de hijos
Primaria	2.27	2
Secundaria	1.66	2
Mayor	0.57	0

Con la tabla 14, se observan diferencias entre los valores de la media y la mediana para las categorías *primaria*, *secundaria* y *mayor*, que pueden estar relacionados con la presencia de datos atípicos. Por lo tanto, se elaboran tres boxplots que representen la información del número de hijos nacidos con respecto a cada una de las tres categorías del nivel educativo (teniendo en cuenta algunos elementos para la elaboración de cada boxplot presentados en la tabla 15).

Figura 10: Boxplots del número de hijos nacidos según los niveles educativos: *primaria*, *secundaria* y *mayor*

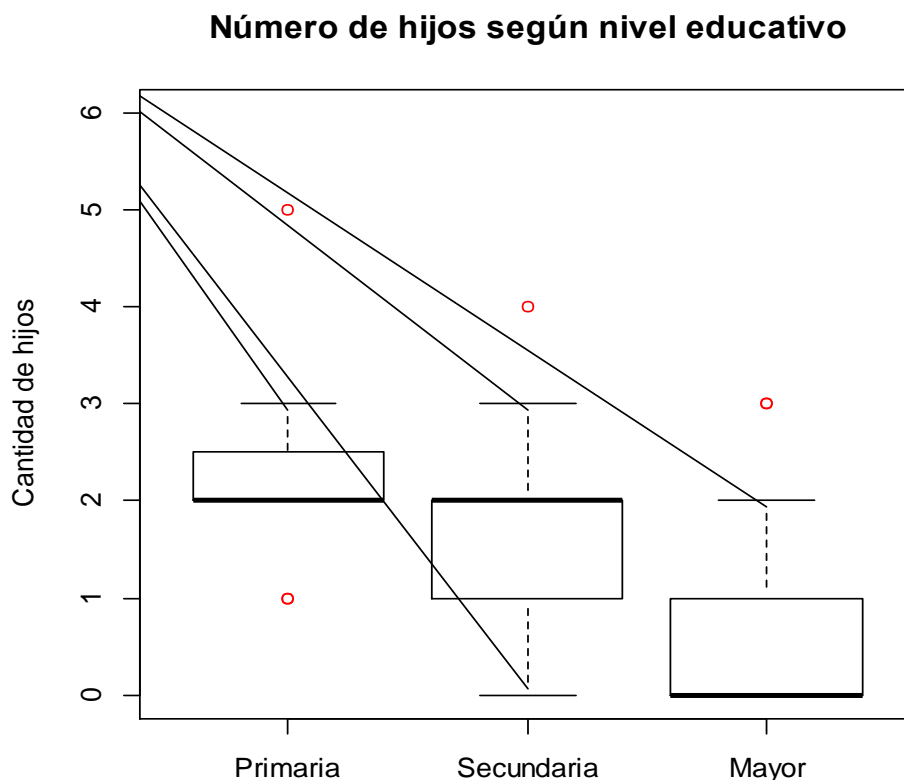


Tabla 15: Elementos para la elaboración de boxplots del número de hijos nacidos según los niveles educativos primaria, secundaria y mayor

Elementos	Primaria	Secundaria	Mayor
Primer cuartil	2	1	0
Segundo cuartil	2	2	0
Tercer cuartil	2.5	2	1
Rango intercuartílico	0.5	1	1
Mínimo	1	0	0
Máximo	5	4	3
Bigote.inf	2	0	0
Bigote.sup	3	3	2

Para este caso, la presencia de datos atípicos en las categorías *primaria*, *secundaria* y *mayor* afectan el cálculo de la media aritmética, de manera que se recomienda como medida resumen, más apropiada, la mediana porque se ve menos afectada por la presencia de estos datos atípicos.

3.3. Utilización de un software estadístico

Como se menciona en la sección 3.2.3, durante el desarrollo de la actividad los estudiantes tienen la posibilidad de utilizar un software estadístico como R-Project o Microsoft Excel. En esta sección se mencionan algunos detalles durante la exploración de estas herramientas tecnológicas.

3.3.1. Utilización de comandos básicos en R-Project

Previamente a la implementación de esta estrategia didáctica, se realizan algunas actividades, con los estudiantes mencionados en la sección 1.1, para el reconocimiento y uso de algunos comandos básicos -referentes a estadística descriptiva- que ofrece el software libre R-Project (Ver Anexo D. Actividad de reconocimiento en R-Project). Adicionalmente, se realiza una actividad para trabajar la importación de datos desde Microsoft Excel a R-Project (Ver Anexo E. Importar datos a R-Project).

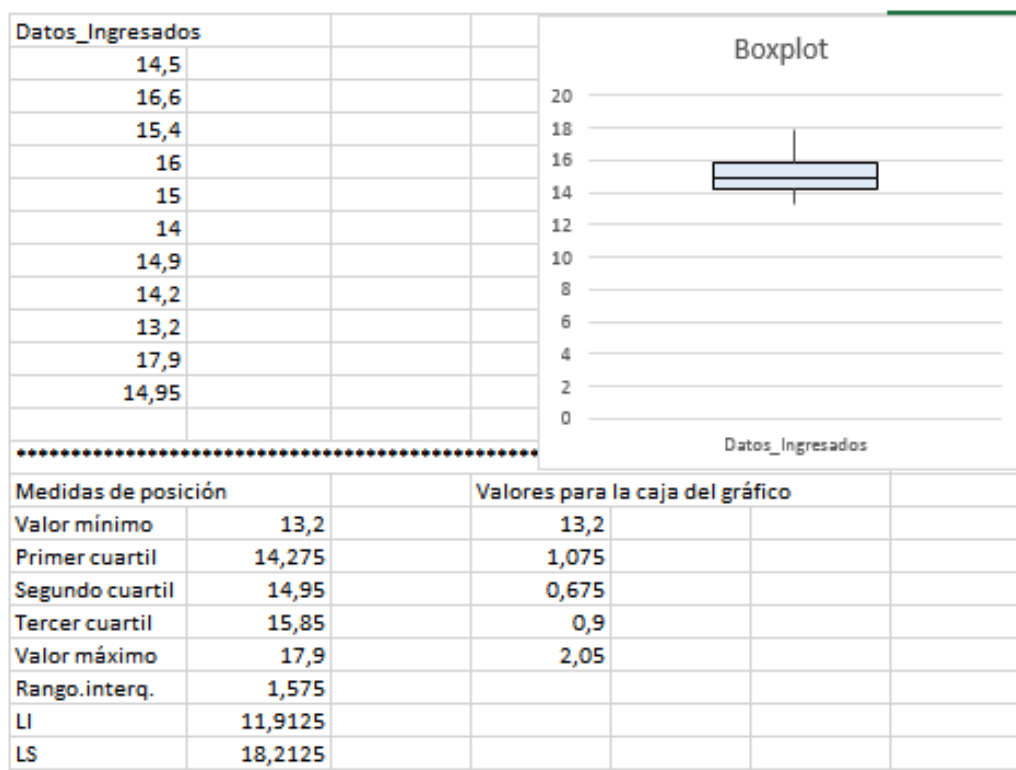
Entretanto, en un ejercicio de comparación entre los (posibles) cálculos manuales de un estudiante y los cálculos presentados por R-Project, se hace énfasis en que para el tratamiento de variables continuas por medio de cálculos manuales se necesitan agrupar

los datos en intervalos y, en consecuencia, se realiza un análisis aproximado donde se pierde información. Por el contrario, el software estadístico utiliza un tratamiento de datos enumerados o enlistados para realizar un análisis exacto.

3.3.2. Exploración de Excel para la elaboración de un boxplot

Teniendo en cuenta que los estudiantes están familiarizados con el software Microsoft Excel (dado que su utilización hace parte del currículo de la institución) y que éste no tiene una plantilla para la elaboración de un boxplot, en esta sección se diseña una macro simple para los cálculos de algunas medidas de posición y la producción de este gráfico. Específicamente, la macro contiene diez entradas de datos (que pueden ser modificados al momento de la exploración), algunos cálculos como las medidas de posición (valor mínimo, primer cuartil, mediana, tercer cuartil, valor máximo, rango intercuartílico, límite inferior y límite superior) y su respectivo boxplot (ver figura 11).

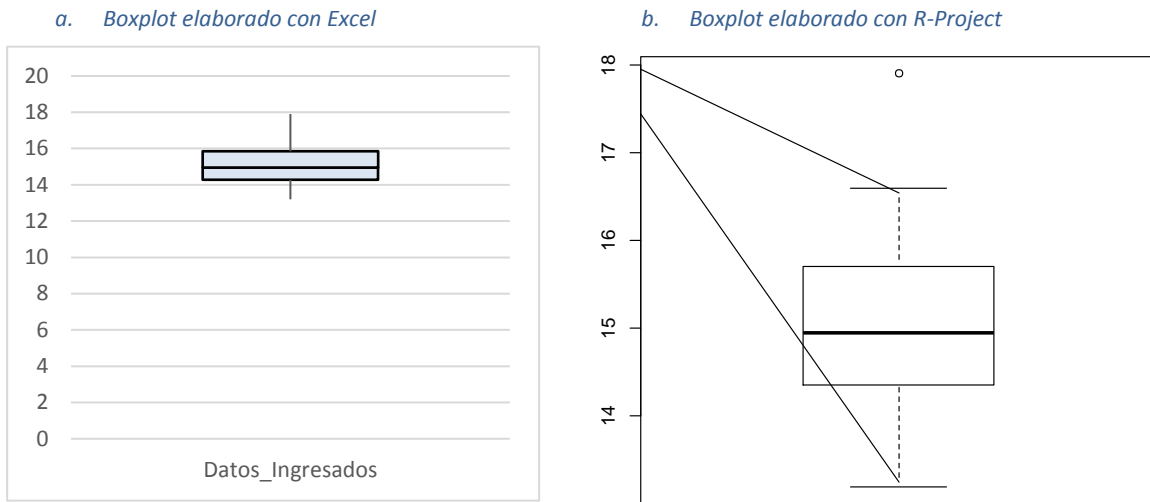
Figura 11: Macro en Excel para elaborar un boxplot



Cabe destacar que en la utilización de este software para la elaboración de un boxplot se presentan algunas dificultades como no poder visualizar los datos atípicos (cuando existen)

o manejar una escala automática que no se ajusta al rango de los datos. En contraste, el gráfico que elabora el software R-Project no presenta estas dificultades y se ajusta a los contenidos disciplinares de este trabajo (Ver figura 12).

Figura 12: Dos boxplot construidos con diferente software para un mismo conjunto de datos



Capítulo 4: Estrategia didáctica: Resultados y análisis

En este capítulo se analiza la información obtenida, con respecto a los resultados o producciones del grupo de estudiantes mencionado en la sección 1.1, tanto en la prueba diagnóstica como en el desarrollo del proyecto de Estadísticas de sexualidad.

4.1. Prueba diagnóstica: análisis de resultados

La prueba diagnóstica se aplica en dos oportunidades con finalidades distintas; en un primer momento, se aplica a 28 estudiantes de grado noveno, en un lapso máximo de 70 minutos, para identificar sus conocimientos previos. Luego, en un segundo momento, después de la implementación del proyecto Estadísticas de sexualidad, se aplica nuevamente la prueba con algunas modificaciones (se le denomina prueba final) y durante el mismo tiempo para valorar los conocimientos adquiridos por la mayoría de este mismo grupo de estudiantes.

Nota: Esta estrategia didáctica se implementa con un grupo de estudiantes descrito en la sección 1.1. Sin embargo, por razones como el retiro de un estudiante de la institución o el traslado de alguno de ellos, no fue posible aplicar todos los instrumentos (las pruebas [diagnóstica y final] y el proyecto Estadísticas de sexualidad) al mismo número de alumnos.

4.1.1. Análisis de las preguntas de opción múltiple

De las diez preguntas realizadas en la prueba diagnóstica, las siete primeras son de selección múltiple y las últimas tres son abiertas. Con los resultados de las primeras siete preguntas durante el primer momento de aplicación (ver tabla 16) se observa que las preguntas 1, 4 y 5 presentan un nivel de acierto por debajo del 36%.

Tabla 16: Resultados de la prueba diagnóstica

Número de pregunta	1	2	3	4	5	6	7
Proporción de respuestas correctas	$\frac{7}{28}$	$\frac{28}{28}$	$\frac{20}{28}$	$\frac{10}{28}$	$\frac{10}{28}$	$\frac{25}{28}$	$\frac{21}{28}$

Al tener en cuenta la opción errada con mayor frecuencia para cada una de estas preguntas se encuentra que posiblemente los estudiantes obvian el punto cero como inicio de una escala de valores, interpretan el valor de la frecuencia absoluta (en una tabla) como el equivalente a su respectivo porcentaje o confunden la frecuencia absoluta con la frecuencia

relativa. Específicamente, se observa que los estudiantes no dominan el concepto de porcentaje y presentan errores cuando realizan las operaciones básicas para sus cálculos.

Entretanto, al estudiar las estructuras de estas preguntas con bajo acierto (4, 5 y 6) se encontraron algunas fallas técnicas como utilizar representaciones gráficas con distintas escalas para las opciones de respuestas correspondientes a la pregunta 1 u omitir información en los enunciados de las preguntas 4 y 8, lo que posiblemente altera la interpretación del estudiante. Por lo tanto, se realiza el respectivo ajuste en la prueba final para el segundo momento de aplicación (Ver anexo E. Modificaciones a la prueba diagnóstica).

Con las respuestas de la prueba final (durante el segundo momento de aplicación), se observa que el número de aciertos para la pregunta 1 sube considerablemente indicando que la mayoría de estudiantes reconocen los convenios de construcción y elementos de un gráfico de barras vertical para los niveles de una variable categórica. Adicionalmente, el número de aciertos aumenta para las preguntas 4 y 5, aunque, específicamente, para la pregunta 4 no se consideran satisfactorios los resultados porque no alcanzan a aprobar el 50% de los estudiantes, de manera que se consigue parcialmente que los estudiantes conozcan y dominen el concepto de porcentaje en diversos contextos (ver tabla 17).

Tabla 17: Resultados de la prueba final

Número de pregunta	1	2	3	4	5	6	7
Proporción de respuestas correctas	$\frac{19}{26}$	$\frac{25}{26}$	$\frac{22}{26}$	$\frac{12}{26}$	$\frac{13}{26}$	$\frac{23}{26}$	$\frac{16}{26}$

4.1.2. Análisis estadístico inferencial de las pruebas: diagnóstica (antes) y final (después)

En este análisis no se cuenta con una muestra aleatoria de la población por cuestiones logísticas. Estos análisis son generalizables a poblaciones similares a las de este estudio (detalles en la sección 1.1).

De los estudiantes del curso, solamente 23 estudiantes presentaron las dos pruebas, antes y después de aplicar el proyecto Estadísticas de sexualidad; es decir, 28 estudiantes presentaron la prueba ANTES y 26 la prueba DESPUÉS, pero sólo 23 de ellos presentaron las dos pruebas. Aquí se analizan los resultados de las 7 preguntas de selección múltiple, cada una con 4 opciones de respuesta; una correcta y tres incorrectas. Se codifican, para

el análisis, como cero la respuesta incorrecta y uno la respuesta correcta. En cada pregunta, se tienen las 46 respuestas, dos por cada estudiante.

Test de McNemar. Este contraste estadístico se diseña para comparar proporciones correlacionadas, como en este caso, teniendo en cuenta que se tienen dos datos del mismo estudiante, antes y después de la intervención (Hoffman, 1976).

Como el tamaño de muestra es pequeño, $n = 23$, el test de McNemar se realiza con la corrección del caso. Para cada pregunta se presentan los datos en una Tabla de contingencia 2x2, como se muestra en la tabla 18.

Tabla 18: Tabulación de datos para el test de McNemar

		DESPUÉS		Total
		0	1	
ANTES	0	a	b	$a + b$
	1	c	d	$c + d$
Total				23

Se pueden definir las siguientes proporciones de interés:

$$P_b = P_{buena} = \frac{b}{a+b} \qquad P_c = P_{mala} = \frac{c}{c+d}$$

Donde P_b representa la probabilidad de pasar de respuesta incorrecta, antes de la intervención, a respuesta correcta, después de la intervención, para la misma pregunta. A su vez, P_c representa la probabilidad de pasar de respuesta correcta, antes de la intervención, a respuesta incorrecta, después de la intervención, para la misma pregunta.

El test de McNemar contrasta las siguientes hipótesis:

$$H_0: b = c \qquad vs \qquad H_a: b \neq c$$

H_0 indica que las discordancias son iguales. Con los datos de los 23 estudiantes (que presentaron las dos pruebas), si la hipótesis nula sale favorecida, esto significa que no hubo ganancia con esta intervención (la aplicación del proyecto Estadísticas de sexualidad). La estadística de prueba, usando corrección por continuidad, está dada por:

$$\frac{(|b-c|-1)^2}{b+c}, \text{ la cual distribuye } \chi_1^2, \text{ bajo } H_0 \text{ cierta.}$$

Usando un error tipo I, $\alpha = 0.05$, se rechaza H_0 si la estadística calculada es mayor que 3.85. Se utiliza esta metodología para cada pregunta.

Conclusiones para cada pregunta.

Pregunta 1.

Tabla 19: Datos de la prueba: pregunta 1

Pregunta 1		DESPUÉS		Total
		0	1	
ANTES	0	4	14	18
	1	2	3	5
Total				23

Estadística calculada: 7.56. Decisión: se rechaza H_0 , las discordancias son diferentes.

Conclusión: Como $P_b = 0.78$ y $P_c = 0.40$, se concluye que la intervención tuvo efecto positivo con los estudiantes, en la pregunta 1.

Con un análisis similar, se resumen las siguientes 6 preguntas.

Pregunta 2. (caso especial): en la aplicación de la prueba ANTES de la intervención, los 23 estudiantes contestaron correctamente. Para la aplicación de la prueba DESPUÉS, solamente un estudiante contesta incorrectamente. En ese caso, $P_b = 0.04$ y P_c no está definida.

Pregunta 3.

Tabla 20: Datos de la prueba: pregunta 3

Pregunta 3		DESPUÉS		Total
		0	1	
ANTES	0	1	6	7
	1	2	14	16
Total				23

Estadística calculada: 1.125. Decisión: No se rechaza H_0 .

Conclusión: No hay evidencia estadística de que las discordancias sean diferentes en la pregunta 3. Aunque $P_b = 0.86$ y $P_c = 0.13$, el efecto positivo de la intervención (pasar de una respuesta incorrecta a una correcta) no fue estadísticamente mejor que el efecto negativo (pasar de una respuesta correcta a una incorrecta) con los estudiantes, en la pregunta 3.

Pregunta 4.

Tabla 21: Datos de la prueba: pregunta 4

Pregunta 4		DESPUÉS		Total
		0	1	
ANTES	0	9	6	15
	1	4	4	8
Total				23

Estadística calculada: 0.10. Decisión: No se rechaza H_0 .

Conclusión: No hay evidencia estadística de que las discordancias sean diferentes en la pregunta 4. En este caso, al tener $P_b = 0.40$ y $P_c = 0.50$, puede interpretarse, de manera descriptiva, que hay más efecto negativo que positivo de la intervención con los estudiantes, en la pregunta 4.

Pregunta 5.

Tabla 22: Datos de la prueba: pregunta 5

Pregunta 5		DESPUÉS		Total
		0	1	
ANTES	0	8	7	15
	1	3	5	8
Total				23

Estadística calculada: 0.90. Decisión: No se rechaza H_0 .

Conclusión: No hay evidencia estadística de que las discordancias sean diferentes en la pregunta 5. En este caso, aunque $P_b = 0.47$ y $P_c = 0.38$, el efecto positivo de la intervención no es estadísticamente mejor que el efecto negativo con los estudiantes, en la pregunta 5.

Pregunta 6.

Tabla 23: Datos de la prueba: pregunta 6

Pregunta 6		DESPUÉS		Total
		0	1	
ANTES	0	0	2	2
	1	3	18	21
Total				23

Estadística calculada: 0.0. Decisión: No se rechaza H_0 .

Conclusión: No hay evidencia estadística de que las discordancias sean diferentes en la pregunta 6. En este caso, aunque $P_b = 1$ y $P_c = 0.14$, el efecto positivo de la intervención no es estadísticamente mejor que el efecto negativo con los estudiantes, en la pregunta 6.

Pregunta 7.

Tabla 24: Datos de la prueba: pregunta 7

Pregunta 7		DESPUÉS		Total
		0	1	
ANTES	0	2	4	8
	1	6	11	17
Total				23

Estadística calculada: 0.10. Decisión: No se rechaza H_0 .

Conclusión: No hay evidencia estadística de que las discordancias sean diferentes en la pregunta 7. En este caso, aunque $P_b = 0.67$ y $P_c = 0.36$, el efecto positivo de la intervención no es estadísticamente mejor que el efecto negativo con los estudiantes, en la pregunta 7.

4.1.3. Análisis cualitativo de las preguntas abiertas

Primer momento de aplicación

Con las respuestas de las preguntas abiertas (8, 9 y 10) se tienen en cuenta las variables estadísticas representadas en el gráfico y otros objetos matemáticos como expresiones simbólicas, procedimientos y conceptos (Batanero, Arteaga & Ruíz, 2009) (que los estudiantes utilizaron para su resolución), para diseñar niveles de respuesta y analizar los gráficos estadísticos producidos con base en la información de la tabla 25.

Tabla 25: Estudiantes que utilizaron elementos estructurales en sus gráficos durante la prueba diagnóstica

Elementos	Descripción	Pregunta		
		8	9	10
Palabras	Título	2	14	3
	Etiquetas	13	18	9
	Rótulos	11	4	4
Marco	Uso de ejes	n.a.	23	11
	Escala proporcional	n.a.	12	5
	Marcas de referencia	20	22	11
	Líneas de división del círculo	20	n.a.	n.a.
Especificadores	Uso de barras	n.a.	22	11
	Uso de sectores circulares	19	n.a.	n.a.
Convenios	Altura de las barras proporcional a la frecuencia	n.a.	14	6
	Amplitud del sector proporcional a la frecuencia	4	n.a.	n.a.
Contenido	Identifica la variable y los datos a representar	7	16	2

n.a.=no aplica

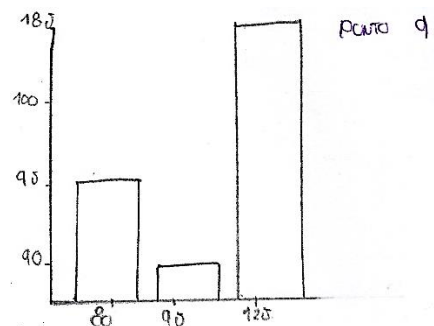
Nivel 0. No realiza el gráfico. Aunque en la prueba se requiere la construcción de un gráfico estadístico para cada uno de los puntos 8, 9 y 10, el estudiante no lo construye posiblemente por no poder completar, interpretar o leer los datos de una tabla de contingencia (para el punto 8); por no identificar la variable y los datos a representar; por no saber aplicar el cálculo de porcentajes (para el punto 10) o por no conocer los convenios de construcción de gráficos como el diagrama de sectores o el diagrama de barras (ver tabla 25).

Nivel 1. No reconoce la información a representar ni utiliza elementos estructurales. Para este nivel de respuesta los estudiantes utilizan información diferente o adicional para

representar su gráfico estadístico, de manera que no se logra identificar la variable y los datos a representar. Además, en su representación gráfica omite algunos elementos estructurales como *un título*, *rótulos* o *convenios* que se consideran fundamentales para la interpretación de los datos.

Los gráficos requeridos para cada una de estas preguntas corresponden a una variable nominal con dos categorías, pero en algunos gráficos producidos por los estudiantes -en este nivel- se incluye el total (la distribución marginal) en su representación, lo que se considera incorrecto porque con el gráfico se busca ilustrar la proporción de encuestados en cada una de las categorías (Ver figura 13).

Figura 13: Ejemplo de un gráfico nivel 1 producido por un estudiante

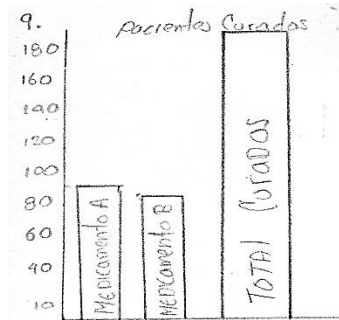


En este nivel de respuesta el estudiante sólo busca relacionar “un elemento de un eje con el de otro eje” (Arteaga, 2008, p.24) sin interpretación de la información representada, de manera que se clasifica como *leer entre los datos* según los niveles de lectura propuestos por Curcio. Además, con base en los niveles de comprensión de gráficos estadísticos propuestos por Gerber, Boulton-Lewis y Bruce, con este nivel de respuesta se observa que los estudiantes no se centran en los datos o los referencian de forma incompleta representando sólo aspectos parciales.

Nivel 2. No reconoce la información a representar. Utiliza información diferente o adicional para representar su gráfico estadístico, de manera que el estudiante no logra identificar la variable y los datos a representar, pero hace uso de algunos elementos estructurales. A diferencia del nivel anterior, el estudiante utiliza algunos elementos como *etiquetas*, *un título*, o *convenios* en la construcción de su gráfico, lo que le permite a un lector reconocer el tipo de información que se quería ilustrar, aunque se haya utilizado datos incorrectos (Ver figura 14).

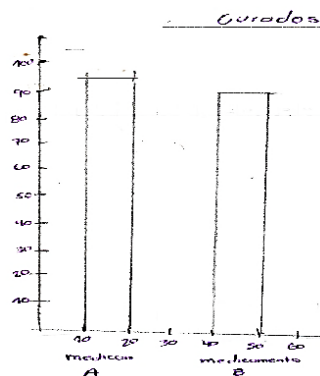
En este nivel de respuesta los estudiantes aun no consiguen interpretar la información contenida en la tabla de contingencia y representarla en el gráfico, a pesar de que utilizan algunos elementos estructurales para indicar cuáles son las variables representadas, de manera que se clasifica como *leer entre los datos*.

Figura 14: Ejemplo de un gráfico nivel 2 producido por un estudiante



Nivel 3. No tiene en cuenta los elementos estructurales de un gráfico. Lee e interpreta los datos de una tabla de contingencia y logra reconocer correctamente la variable y la información que necesita representar, pero omite elementos estructurales en la construcción del gráfico como *etiquetas, rótulos, un título, marcas de referencia, convenios* o uso de porcentajes, lo que altera la interpretación de la información ilustrada por parte de un lector externo (Ver figura 15).

Figura 15: Ejemplo de un gráfico nivel 3 producido por un estudiante



En este nivel de respuesta, se alcanza el nivel de comprensión *leer dentro de los datos* porque el estudiante representa en su gráfico la relación entre dos subconjuntos de datos permitiéndole interpretar e integrar la información (Arteaga, 2008), aunque no se utilicen algunos elementos estructurales. Paralelamente, se consigue llegar al tercer nivel de comprensión que proponen Gerber, Boulton-Lewis y Bruce dado que se aprecia el propósito

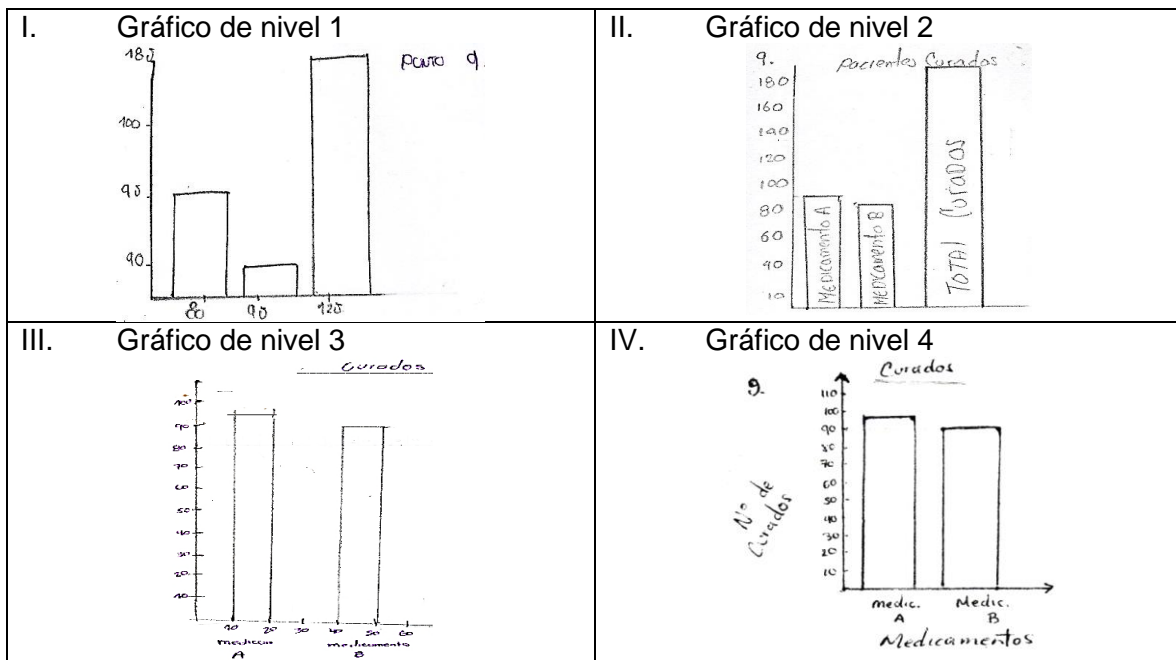
del gráfico (aunque no es posible obtener una síntesis global por la omisión o la falta de comprensión de un elemento clave en la representación).

Nivel 4. Construye correctamente un gráfico utilizando sus elementos estructurales. Además de reconocer la variable y los datos a representar, utilizan elementos estructurales para la construcción de su gráfico evitando interpretaciones erradas a causa de falta de información. Es decir, en el gráfico se identifica la variable categórica por medio del rótulo, las categorías de esta variable por medio de las etiquetas, las marcas de referencia y la etiqueta del *eje vertical* para ilustrar la cantidad de encuestados en cada una de estas categorías (para las preguntas 9 y 10) (Ver figura 16).

Este nivel de respuesta se clasifica en *leer dentro de los datos* y, a diferencia del nivel de respuesta anterior, permite llegar a una síntesis global. Cabe mencionar que (para esta prueba) a los estudiantes no se les pide realizar predicciones o inferencias de los datos, por lo tanto, no se tienen en cuenta otros niveles para la lectura o comprensión de gráficos estadísticos.

En la figura 16 se muestran algunos ejemplos para cada nivel de respuesta tomando como referencia el punto 9 de la prueba diagnóstica (Ver anexo A. Prueba diagnóstica).

Figura 16: Ejemplos de gráficos producidos por estudiantes según el nivel de respuesta



Además, en la tabla 26 se presenta la clasificación de los estudiantes respecto a los niveles de respuesta propuestos, teniendo en cuenta las preguntas 8, 9 y 10 de la prueba diagnóstica.

Tabla 26: Clasificación de estudiantes según el nivel de respuesta durante la prueba diagnóstica

Nivel de respuesta	Pregunta		
	8	9	10
0. No realiza el gráfico	5	7	14
1. No reconoce la información a representar ni los elementos estructurales	8	2	2
2. No reconoce la información a representar	11	6	5
3. No tiene en cuenta los elementos estructurales	2	12	7
4. Construye correctamente el gráfico	2	1	-

Segundo momento de aplicación

Con base en las respuestas de las preguntas 8, 9 y 10 de la prueba final durante el segundo momento de aplicación, se resumen en la tabla 27 la cantidad de elementos estructurales que utilizaron los estudiantes en la construcción de sus gráficos estadísticos.

Tabla 27: Estudiantes que utilizaron elementos estructurales en sus gráficos durante la prueba final

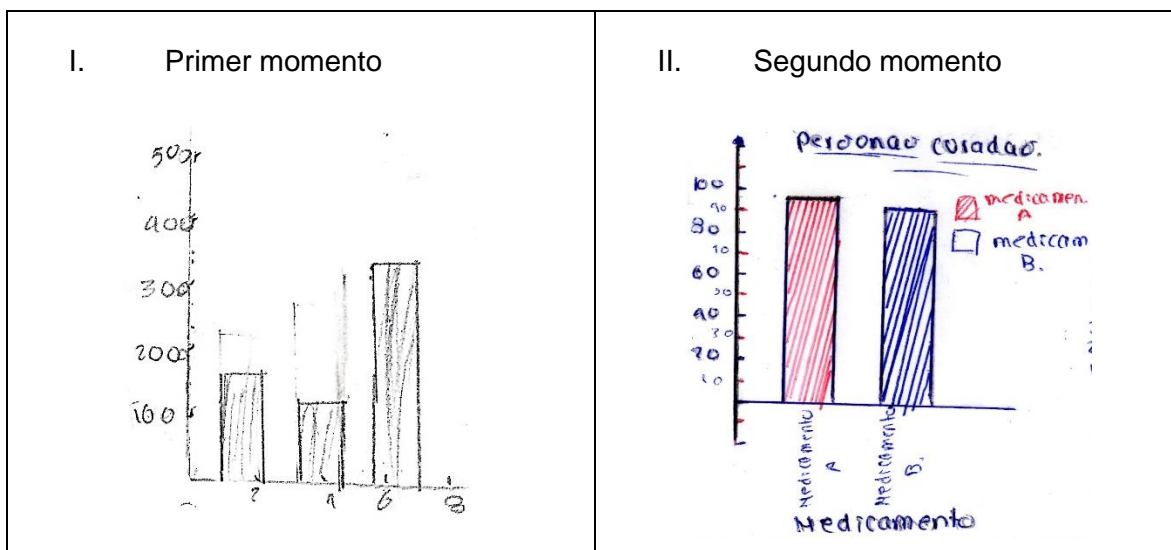
Elementos	Descripción	Pregunta		
		8	9	10
Palabras	Título	5	17	13
	Etiquetas	21	23	17
	Rótulos	17	12	7
Marco	Uso de ejes	n.a.	24	18
	Escala proporcional	n.a.	21	15
	Marcas de referencia	21	23	18
	Líneas de división del círculo	23	n.a.	n.a.
Especificadores	Uso de barras	n.a.	23	18
	Uso de sectores circulares	22	n.a.	n.a.
Convenios	Altura de las barras proporcional a la frecuencia	n.a.	17	8
	Amplitud del sector proporcional a la frecuencia	12	n.a.	n.a.
Contenido	Identifica la variable y los datos a representar	18	19	14

n.a.=no aplica

Comparando los datos de las tablas 25 y 27, se puede observar que la cantidad de estudiantes que utilizan elementos estructurales en sus gráficos estadísticos aumenta en cada una de las categorías con respecto al primer momento de aplicación de la prueba diagnóstica. Específicamente, la cantidad de estudiantes que *identifican la variable y los datos a representar* aumenta considerablemente y les permite avanzar a los niveles de respuesta tres y cuatro.

Por ejemplo, en la figura 17 se observa la evolución de un gráfico estadístico elaborado por uno de los estudiantes durante la aplicación de las pruebas (para contestar la pregunta 9), dado que en el primer momento de aplicación no reconoce los datos que necesitaba representar, las *palabras* informativas del gráfico (como un título, etiquetas y rótulos) y algunos convenios en la construcción de un gráfico de barras como la asignación de categorías (para diferenciar la información de cada barra) a uno de los ejes (Ver figura 17.I). Luego, en el segundo momento de aplicación no solamente logra reconocer la variable y los datos que necesita representar, sino que utiliza los elementos estructurales para un gráfico de barras como un título, etiquetas, un rótulo y los convenios de construcción como la altura proporcional de las barras con respecto a su frecuencia (Ver figura 17.II).

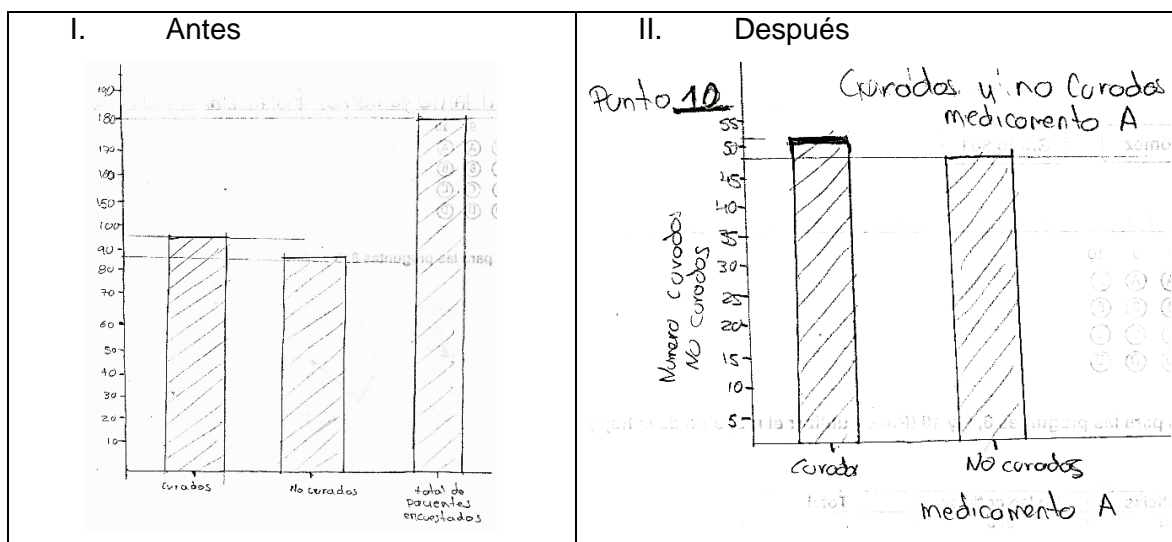
Figura 17: Gráficos producidos por un estudiante para responder la pregunta 9 de la prueba en los dos momentos de aplicación



Adicionalmente, en la figura 18 se muestra las producciones de otro estudiante para responder a la pregunta 10 y se observa que, durante el primer momento de aplicación, no reconoce los datos que necesita representar, no utiliza una escala proporcional para el eje vertical, posiblemente no puede realizar conversiones de frecuencias a términos

porcentuales y no utiliza algunos elementos estructurales como un título o etiquetas (Ver figura 18.I). Luego, en el segundo momento logra construir un gráfico reconociendo los datos a representar y expresándolos por medio de cantidades porcentuales, teniendo en cuenta los diversos elementos estructurales para un gráfico de barras. Cabe mencionar que la etiqueta utilizada para el eje vertical presenta un error porque las cantidades están en términos porcentuales (ver figura 18.II).

Figura 18: Gráficos producidos por un estudiante para responder la pregunta 10 de la prueba en los dos momentos de aplicación



Finalmente, se elabora la tabla 28 para clasificar a los estudiantes con respecto al nivel de respuesta, encontrando que, a diferencia del primer momento de aplicación (ver tabla 26), por lo menos la mitad de los estudiantes se clasifican en los niveles 3 y 4 con respecto a cada una de las preguntas abiertas (8, 9 y 10). Además, la cantidad de estudiantes que no realizan un gráfico estadístico -Nivel 0- baja considerablemente (por ejemplo, en la pregunta 9 la cifra de estudiantes que no elaboran el gráfico baja de 7 a 2 entre los dos momentos de aplicación).

Tabla 28: Clasificación de estudiantes según el nivel de respuesta durante la prueba final

Nivel de respuesta	Pregunta		
	8	9	10
0. No realiza el gráfico	3	2	7
1. No reconoce la información a representar ni los elementos estructurales	1	1	1
2. No reconoce la información a representar	5	2	5
3. No tiene en cuenta los elementos estructurales	13	8	9
4. Construye correctamente el gráfico	4	13	4

4.2. Proyecto Estadísticas de sexualidad: análisis de resultados

Con el desarrollo del proyecto *Estadísticas de sexualidad* se evalúan los conocimientos aprendidos por los estudiantes, durante las clases de estadística descriptiva, y la capacidad de aplicarlos en contexto (haciendo énfasis en la construcción de gráficos estadísticos como un cambio de registro de representación).

Para el desarrollo del proyecto se utilizan cuatro sesiones presenciales en el aula de clase, tres sesiones de 1 hora y 40 minutos y una sesión de 50 minutos. En la primera sesión se conforman grupos de cuatro integrantes, se entrega el material (ver sección 3.2.3) y se brindan algunas instrucciones generales; en esta sesión no hubo acceso a computadores. Para las siguientes tres sesiones, la actividad se realiza en la sala RED-P, sala de cómputo, donde los estudiantes tienen la posibilidad de trabajar el proyecto utilizando cálculos manuales o cálculos por medio del software R-Project.

Entre los estudiantes que utilizan este software, se observa que algunos de ellos ingresan equivocadamente algunos comandos a R-Project por falta de especificación como no reconocer previamente el encabezado de los datos o no reconocer la variable de interés (para determinada pregunta). Estas dificultades fueron resueltas oportunamente con la orientación del docente.

Luego de las cuatro sesiones presenciales, cada grupo de estudiantes entrega un informe escrito teniendo en cuenta los resultados y análisis del proyecto desarrollado. Sin embargo, en algunos informes grupales se evidencia un trabajo aislado de cada integrante del grupo con respecto a una tarea particular del proyecto; por lo tanto, se solicita la entrega de un informe individual para conocer los resultados del proyecto por parte de cada estudiante.

Análisis de los gráficos estadísticos producidos por los estudiantes

Teniendo en cuenta que los primeros tres puntos de cada proyecto son iguales para todos los grupos (ver sección 3.2.3) y que los estudiantes tienen la libertad de utilizar cálculos, tablas o gráficos que consideren pertinentes, se realiza la tabla 29 para brindar información referente a la producción de gráficos estadísticos por parte de este grupo de alumnos.

Nota: Del total de estudiantes que trabajan en el proyecto, solamente 18 de ellos entregan el informe escrito (individual) en el tiempo establecido. Cabe resaltar que no todos los

estudiantes utilizan gráficos estadísticos para resolver una determinada pregunta de su proyecto (por ejemplo, en la tabla 29 se observa que, de los 18 estudiantes, solamente 7 de ellos utilizan un gráfico estadístico para responder la pregunta 1). Con base en estos informes individuales se realiza el análisis.

Tabla 29: Información acerca de los gráficos producidos por los estudiantes para las primeras tres preguntas del proyecto

Aspectos considerados	Pregunta		
	1	2	3
Realiza gráfico	7	7	17
No realiza gráfico	11	11	1
Total	18	18	18
Gráfico correcto	2	1	4
Gráfico parcialmente correcto	2	4	9
Gráfico incorrecto	3	2	4
Total	7	7	17
Gráfico pertinente	5	5	17
Gráfico no pertinente	2	2	-
Total	7	7	17

Para analizar los gráficos estadísticos producidos por los estudiantes se tienen en cuenta algunos niveles de respuesta propuestos en el análisis cualitativo de la prueba diagnóstica (ver sección 4.1.3) y la utilización de elementos estructurales para la elaboración de cada gráfico (Ver tabla 30).

Tabla 30: Estudiantes que utilizaron elementos estructurales en los gráficos presentados en el proyecto

Elementos	Descripción	Pregunta		
		1	2	3
Palabras	Título	5	2	5
	Etiquetas	5	5	3
	Rótulos	3	1	n.a.
Marco	Uso de ejes	7	5	13
	Escala proporcional	6	5	11
	Marcas de referencia	7	5	15
Especificadores	Uso de barras	5	5	-
	Caja y bigotes (Boxplot)	n.a.	n.a.	15
	Línea referente a la mediana (Boxplot)	n.a.	n.a.	12
	Círculos para datos atípicos (Boxplot)	n.a.	n.a.	15
Convenios	Altura de las barras proporcional a la frecuencia	4	5	-
Contenido	Identifica la variable y los datos a representar	7	7	17

n.a.=no aplica

La clasificación de los estudiantes que producen un gráfico estadístico para responder las primeras tres preguntas del proyecto, según el nivel de respuesta, se muestra en la tabla 31.

Tabla 31: Clasificación de estudiantes según el nivel de respuesta en la construcción de gráficos estadísticos

Nivel de respuesta	Pregunta		
	1	2	3
1. No reconoce la información a representar ni los elementos estructurales	-	-	-
2. No reconoce la información a representar	-	-	-
3. No tiene en cuenta los elementos estructurales	5	6	13
4. Construye correctamente el gráfico	2	1	4

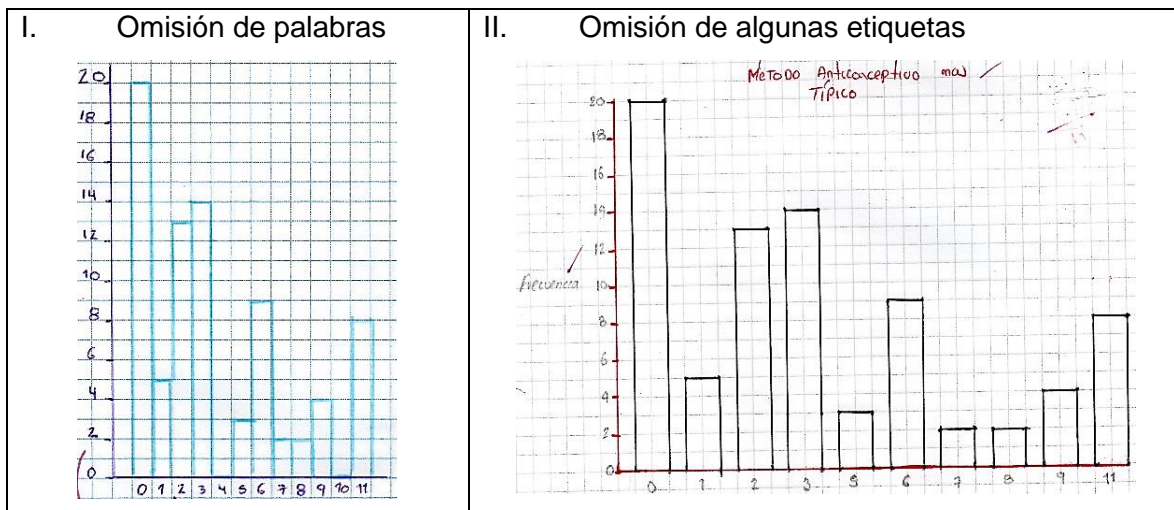
Con las tablas 30 y 31 se puede observar que la totalidad de estudiantes logran reconocer las variables de estudio y los datos a representar para cada una de las tres primeras preguntas del proyecto. Sin embargo, al construir los gráficos estadísticos emergen algunas dificultades por omitir algunos elementos estructurales que alteran la interpretación y análisis de la información representada. A continuación, se describen algunos de los errores encontrados en la construcción de estos gráficos.

Pregunta 1. ¿Cuál es el método anticonceptivo típico utilizado por las mujeres?

En los gráficos producidos por los estudiantes para el desarrollo de la primera pregunta del proyecto se encuentran algunos errores como omitir un título, las etiquetas o algunos rótulos. Por ejemplo, en la figura 19.I se observa que el estudiante representa los datos por medio de un diagrama de barras, pero no le asigna un título, no utiliza etiquetas para cada uno de los ejes ni utiliza rótulos para el eje horizontal. Mientras que en la figura 19.II el estudiante construye un diagrama de barras con un título y una etiqueta para el eje vertical, pero no utiliza un rótulo para el eje horizontal ni hay claridad con respecto a sus etiquetas. Estos dos ejemplos de gráficos se clasifican en el nivel de respuesta tres porque la omisión de estos elementos altera la interpretación de la información representada.

Cabe mencionar que el diagrama de barras utilizado en la figura 19.I presenta un error de diseño porque al representar una variable cualitativa nominal no se utilizan barras adosadas. Adicionalmente, el título del gráfico 19.II es inadecuado porque no sólo se representa el “método anticonceptivo más típico”, sino también todos los métodos anticonceptivos utilizados por las mujeres entrevistadas.

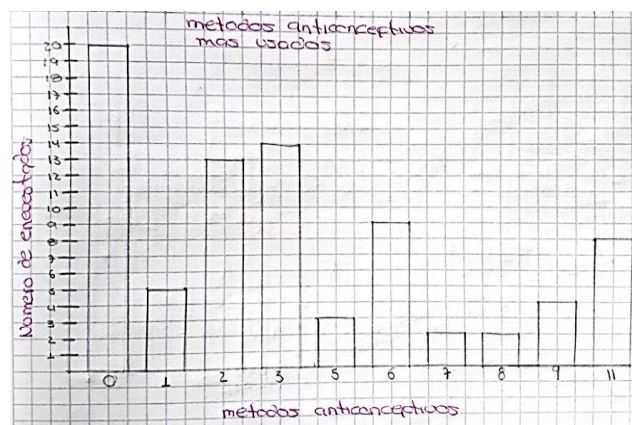
Figura 19: Algunos errores en los gráficos producidos por estudiantes para el desarrollo de la primera pregunta



Por otro lado, los estudiantes que construyen un gráfico correcto -nivel 4- tienen en cuenta un título (que en ocasiones se puede mejorar para poder situar a un lector externo con respecto a la lectura de los datos), las etiquetas de los dos ejes y un rótulo para indicar la variable nominal utilizada. Por ejemplo, en la figura 20 se observa un diagrama de barras con una escala proporcional utilizada para el eje vertical con sus respectivas marcas de referencia y la altura de cada barra es proporcional a la frecuencia absoluta de cada categoría. Sin embargo, el título que se utiliza para este gráfico es incorrecto porque solamente hace referencia a los métodos anticonceptivos más usados y no a todos ellos.

Adicionalmente, la totalidad de estudiantes identifican e interpretan la moda de los datos como la categoría con mayor frecuencia o la barra con mayor altura del gráfico (En este caso corresponde a la categoría “No usa”). Algunos de ellos complementan su respuesta indicando que el método anticonceptivo típico es la “inyección”.

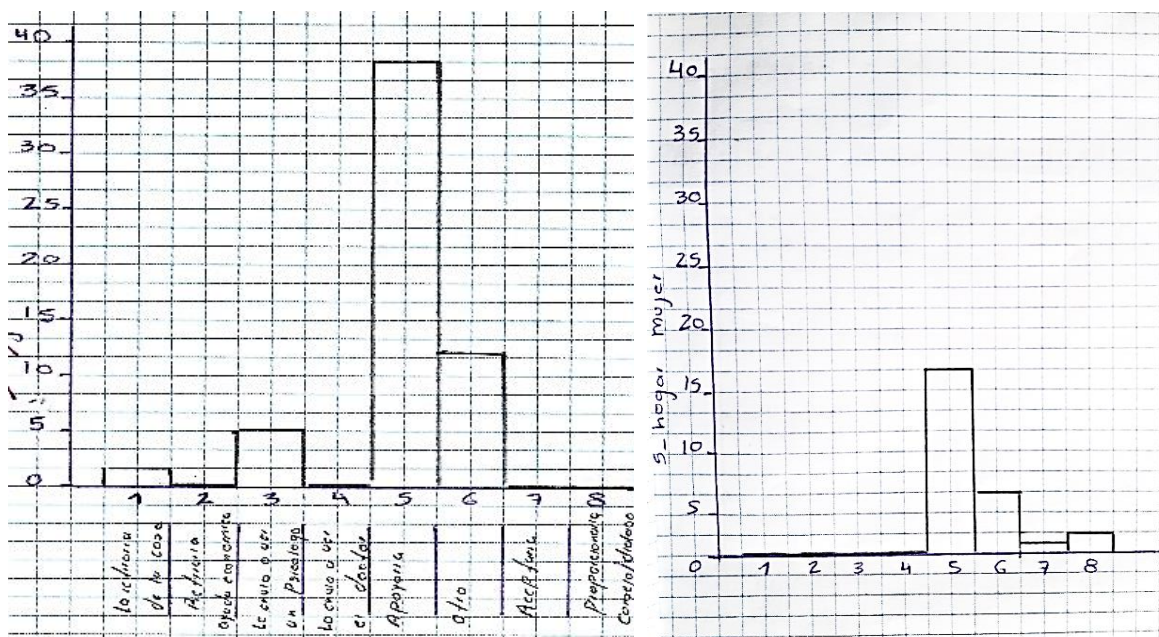
Figura 20: Ejemplo de un gráfico nivel 4 para la primera pregunta



Pregunta 2. ¿Existe alguna relación entre el sexo biológico de la cabeza de hogar con lo que haría si se descubre que uno de sus hijos es homosexual?

De manera análoga a la pregunta anterior, la mayoría de los errores observados se derivan por la falta de *palabras* como elementos estructurales en los gráficos. Sin embargo, también se observaron algunos detalles técnicos que se pueden corregir para mejorar la interpretación del gráfico como evitar construir dos gráficos estadísticos (Ver figura 21) cuando se puede representar la información en uno sólo.

Figura 21: Ejemplo de un gráfico nivel 3 para la segunda pregunta



En la mayoría de las conclusiones, los estudiantes indican cuál es la reacción típica si se descubre que uno de los hijos es homosexual para hogares con *jefatura masculina* o *jefatura femenina*; es decir, contestan que “la respuesta más escogida es [lo] apoyaría”.

Pregunta 3. ¿Existen valores atípicos en la altura de las mujeres entrevistadas? Si los hay, ¿son valores atípicos pequeños o grandes?

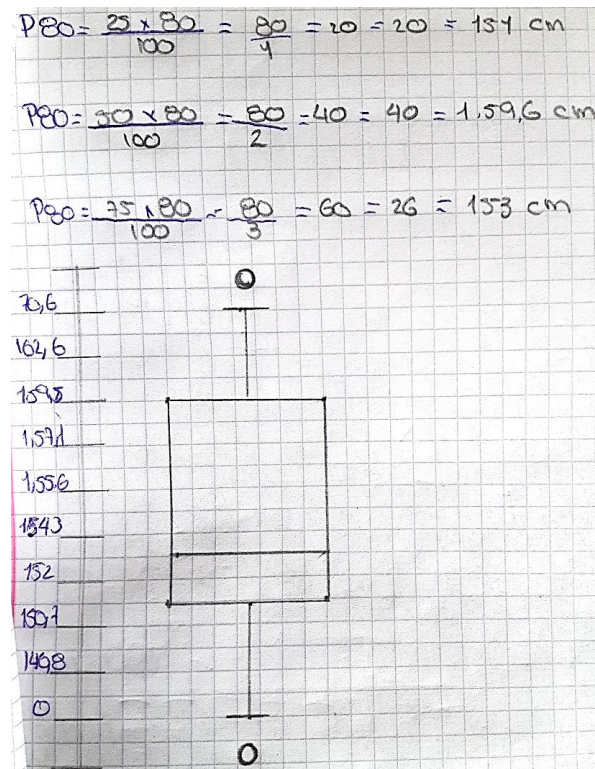
Para responder la pregunta todos los estudiantes, que elaboran un gráfico estadístico, eligen el boxplot para representar la información. Algunos estudiantes utilizan el software R-Project en la elaboración de un boxplot para identificar datos atípicos respecto a la altura de las mujeres entrevistadas. En la mayoría de estos gráficos se omitieron elementos estructurales como un título pertinente o una etiqueta referente al eje y para identificar la cantidad de mujeres encuestadas. Aquellos estudiantes que deciden elaborar el gráfico con operaciones manuales presentan errores en los cálculos de algunos percentiles, límite

superior o inferior (además del cálculo de bigotes) o en utilizar una escala no proporcional para el eje y.

Por ejemplo, en las operaciones para el cálculo de percentiles que se muestran en la figura 22 se observa que la estudiante no utiliza una correcta notación para diferenciar los percentiles 25, 50 y 75, respectivamente. Además, el “percentil 75” resulta menor que el “percentil 50”, lo cual es contradictorio con la definición de percentil. Este tipo de errores en los cálculos conllevan a elaborar incorrectamente un boxplot. Por último, se observa que en el eje y (de su boxplot) asigna valores arbitrarios que no satisfacen la condición de proporcionalidad.

Aun así, se puede observar que en todos los gráficos aparecen dos datos atípicos, uno pequeño y uno grande (posiblemente porque elaboraron el gráfico con base al boxplot que presenta el software R). Algunos de ellos identifican los datos atípico a partir del valor máximo y el valor mínimo de los datos.

Figura 22: Ejemplo de un gráfico nivel 3 para la tercera pregunta



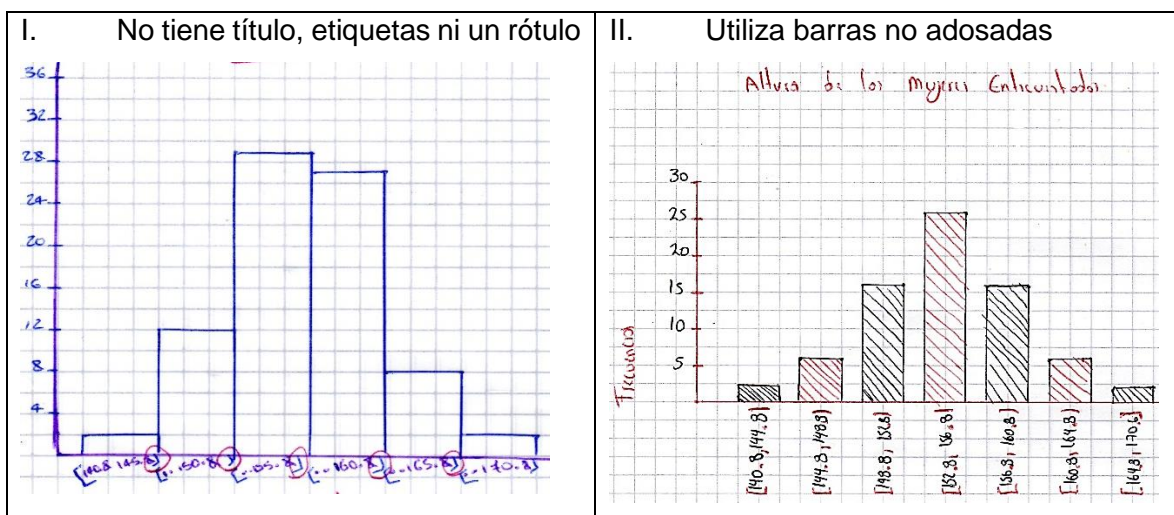
Pregunta 4. ¿Cómo representar la altura de las mujeres entrevistadas?

Para representar la información de la altura de las mujeres entrevistadas los estudiantes utilizan una tabla de distribución de frecuencias o un histograma. Cabe mencionar que para construir el histograma es necesario realizar, primero, una tabla de distribución de frecuencias (si los cálculos son manuales) y se pudo observar que, teniendo en cuenta la naturaleza cuantitativa continua de la variable, para la construcción de las clases o intervalos la mayoría de estudiantes calculan el rango de la variable, el número de intervalos (por el método de Sturges) y su amplitud. Pero algunos de ellos se equivocan al asignar una amplitud (significativamente) diferente al último intervalo con respecto a los demás.

Luego, al estudiar los gráficos producidos se encuentran algunos errores de construcción como omitir elementos estructurales (título, etiquetas o rótulos) o utilizar barras no adosadas (o apiladas) para representar la naturaleza continua de la variable (Ver figura 23).

Se debe agregar que en algunas conclusiones producidas por los estudiantes, se observa que no sólo representan la información en un histograma, sino que además logran extraer información de este gráfico como, por ejemplo, la identificación del intervalo modal. Adicionalmente, algunos de ellos fueron críticos con respecto a la pertinencia de esta pregunta del proyecto *Estadísticas de sexualidad* manifestando que “mientras se hacía el estudio de la altura, [los estudiantes] se preguntaron el por qué se quería saber la altura de las mujeres entrevistadas”.

Figura 23: Ejemplos de algunos errores encontrados en la construcción de un histograma



Pregunta 5. ¿Cuál es el intervalo del peso más común entre las mujeres entrevistadas?

Los resultados de esta pregunta muestran aspectos similares con respecto a la pregunta anterior dado que la mayoría de estudiantes deciden construir una tabla de distribución de frecuencias y, en algunos casos, un gráfico estadístico. Se encontraron errores semejantes al punto anterior como utilizar barras no adosadas para el histograma.

La mayoría de estudiantes contestan de manera correcta la pregunta identificando el intervalo de peso común entre las mujeres entrevistadas a partir del intervalo con mayor frecuencia o la barra con mayor altura. Por lo tanto, los estudiantes que construyeron un histograma alcanzaron el nivel de lectura *Extracción de tendencias* (propuesto por Bertin) porque son capaces de percibir en su gráfico la relación entre dos subconjuntos de datos, es decir, determinan visualmente la moda al clasificar los datos en subconjuntos y luego identificar entre ellos aquel que tiene mayor frecuencia (Arteaga, 2008).

Pregunta: 6. ¿Podrías representar en un gráfico la información referente al conocimiento del periodo fértil en el ciclo ovulatorio utilizando términos porcentuales? Y si ahora restringimos la información entre las mujeres con nivel educativo secundaria, ¿podrías realizar el gráfico?

Esta pregunta tiene dos partes, en la primera se busca representar la información referente al *conocimiento del periodo fértil en el ciclo ovulatorio* utilizando términos porcentuales y en la segunda restringir esta información con el *nivel educativo* secundaria. Para la primera parte, algunos estudiantes elaboraron un gráfico de barras para representar la frecuencia absoluta referente al conocimiento del periodo fértil en el ciclo ovulatorio sin tener en cuenta la instrucción de utilizar términos porcentuales. Luego, en la segunda parte los estudiantes no logran relacionar la información entre las dos variables estadísticas teniendo en cuenta que se construyen gráficos diferentes para representar la frecuencia absoluta de cada una de las variables. De manera que es posible que los estudiantes no logran comprender el contenido de la pregunta, relacionar la información de dos variables categóricas por medio de una tabla de contingencia (pasar de un análisis univariado a un análisis bivariado) o se les dificulta realizar cálculos como los porcentajes.

Pregunta 7. Calcular las medidas resumen del número hijos con respecto a cada nivel educativo, usando dos medidas resumen diferentes: la media y la mediana. ¿Qué diferencias observas entre las dos medidas resumen? ¿Elegirías la media o la mediana para representar la medida resumen del número hijos, según su nivel educativo?

Los estudiantes no logran responder esta pregunta posiblemente por no comprender su contenido. Se considera que esta pregunta presenta un mayor nivel de dificultad en los estudiantes teniendo en cuenta que no sólo tienen que relacionar la información de las variables *número de hijos* y el *nivel educativo*, sino que además deben restringir la información para cada uno de los niveles educativos, realizar los cálculos para determinar la media y la mediana y, finalmente, elegir la medida resumen más apropiada.

Los estudiantes saben calcular la media y la mediana de una variable cuantitativa discreta como *el número de hijos*. Sin embargo, como la información se debe clasificar teniendo en cuenta la variable *nivel educativo* en cuatro subconjuntos de datos (de acuerdo con el número de categorías), posiblemente a los estudiantes se les dificulta relacionar la información de estas dos variables e identificar cada subconjunto de datos para encontrar los valores de la media y la mediana.

Capítulo 5: Conclusiones y recomendaciones

Luego del desarrollo de este trabajo de investigación y del análisis de sus resultados respecto a la aplicación de la estrategia didáctica con énfasis en la producción de gráficos estadísticos, se redactan en este capítulo las conclusiones respecto a cada uno de los objetivos y las recomendaciones para aquellos educadores que deseen poner en práctica esta estrategia didáctica.

5.1. Conclusiones sobre los objetivos

A continuación, se presentan las conclusiones con respecto a cada uno de los objetivos propuestos en el capítulo 1 de esta investigación:

- *Objetivo General. Desarrollar una estrategia didáctica para trabajar con los estudiantes del ciclo IV, la construcción, lectura, interpretación y análisis de gráficos estadísticos mediante la metodología del trabajo por proyectos.*

Este objetivo se cumple con el diseño de la estrategia didáctica descrita en el capítulo 3. En esta estrategia didáctica se elabora un proyecto titulado *Estadísticas de sexualidad* utilizando la metodología del trabajo con proyectos para trabajar la construcción, lectura, interpretación y análisis de gráficos estadísticos con estudiantes de grado noveno, utilizando una sub base (de datos) con 80 registros referentes a temas de planificación familiar y percepción de la educación sexual.

Además, con la metodología del trabajo con proyectos se logra desarrollar una estrategia didáctica pertinente para la construcción, lectura, interpretación y análisis de gráficos estadísticos porque los estudiantes aplican el conocimiento de la estadística descriptiva (específicamente la representación de los datos) para resolver problemas de su interés (como el tema de sexualidad).

- *Objetivo específico. Identificar conocimientos previos de los estudiantes respecto a la construcción, lectura, análisis y uso de información que aparece en gráficos estadísticos.*

Para la recolección de datos acerca de los conocimientos previos de los estudiantes respecto a la construcción, lectura, interpretación y análisis de gráficos estadísticos se

elabora una prueba diagnóstica con siete preguntas de selección múltiple y tres preguntas abiertas.

Se identificaron las siguientes dificultades: obviar el punto cero como inicio de una escala de valores (punto 1) e interpretar el valor de la frecuencia absoluta como el equivalente a su respectivo porcentaje (Punto 4 y 5). Se concluye que la mayoría de estudiantes no tienen en cuenta algunos convenios de construcción de un diagrama de barras y se les dificultan los cálculos de porcentajes.

Además, algunos estudiantes no realizan el gráfico o lo construyen de forma incorrecta posiblemente por no identificar la variable y los datos que necesitaban representar con base en la información representada en una tabla de contingencia. Con lo anterior se puede concluir que la producción de estos gráficos se dificulta por falta de conocimiento de conceptos preliminares como la lectura e interpretación de tablas estadísticas.

- *Objetivo específico. Determinar aspectos epistemológicos, disciplinares y didácticos relacionados con la lectura, interpretación y análisis de gráficos estadísticos.*

La evolución de los gráficos estadísticos durante el último siglo produjo gráficos cada vez más complejos que generaron dificultades en su comprensión, de modo que algunos autores decidieron trabajar en la construcción de niveles de lectura o complejidad de estos gráficos. Incluso, algunos de los errores descritos en investigaciones preliminares, como olvidar utilizar rótulos o utilizarlos de forma equivocada, se presentan en algunos gráficos producidos por los estudiantes durante el desarrollo de este trabajo.

Por otra parte, los conceptos preliminares utilizados para la construcción de un gráfico estadístico varían dependiendo del tipo de gráfico que se utilice. Por ejemplo, a diferencia de un histograma, en la elaboración de un Boxplot se utilizan medidas de posición (entre otras) para la identificación de datos atípicos. Además, el tipo de análisis que se utilice en un estudio puede incidir en la complejidad de la construcción de un gráfico. Por ejemplo, un diagrama de barras se puede utilizar para representar solamente los datos de una variable (análisis univariado) o para representar los datos de dos variables a partir de la lectura e interpretación de una tabla de contingencia (análisis bivariado).

Para finalizar, con la metodología del trabajo con proyectos se puede contextualizar la estadística, hacerla relevante y despertar el interés del estudiante hacia un problema (o tema) de investigación como el tópico de sexualidad abordado en este trabajo.

- *Objetivo específico. Seleccionar situaciones, contextos e información pertinente para estructurar proyectos y diseñar actividades de la estrategia didáctica.*

Teniendo en cuenta que los temas de educación sexual despiertan interés en los estudiantes, se diseña un proyecto, utilizando datos correspondientes a la Encuesta Nacional de Demografía y Salud (2010), para abordar la construcción, lectura e interpretación de gráficos estadísticos. De un total de 1179 variables en el estudio se seleccionaron sólo 16 de ellas teniendo en cuenta la pertinencia en el desarrollo del proyecto dentro del aula de clase y la elección de dos o tres variables por cada clasificación (nominal, ordinal, discreta y continua).

Los datos seleccionados se consideran pertinentes porque se abordan temas de la estadística descriptiva (con respecto a este nivel de secundaria) para organizar y representar información de diferentes tipos de variables estadísticas. Adicionalmente, para contestar algunas preguntas del proyecto es necesario relacionar 2 de las 16 variables presentadas, de manera que el estudiante necesita tratar con información estadística para seleccionar, interpretar y evaluar críticamente información.

- *Objetivo específico. Proponer ejercicios que conduzcan al diseño de macros simples en la hoja de cálculo de Excel para construir gráficos sencillos (opcionalmente, se enseñará la construcción de algunos gráficos en R-Project [software estadístico gratis]).*

El software Microsoft Excel es utilizado regularmente por los estudiantes para el diseño de diagramas de barras. Entre las diversas opciones de gráficos, este software no tiene una plantilla para la elaboración de un boxplot, de manera que se decide realizar una macro simple en una hoja de cálculo que permita calcular algunas medidas de posición y paralelamente construir este tipo de gráfico. Se concluye que el boxplot presentado por Excel no es adecuado para los estudiantes porque no permite visualizar datos atípicos y maneja una escala que no se ajusta al rango de los datos. Además, al momento de explorar otros gráficos que ofrece Excel para las variables cuantitativas continuas se observa que el histograma presentado por este software presenta errores técnicos como la utilización de barras no adosadas para representar la continuidad de una variable.

Por lo tanto, se da prioridad al uso del software libre R-Project en el aula de clase por medio de actividades de reconocimiento de comandos y de introducción al análisis descriptivo a partir de la importación de datos desde un libro de Excel.

- *Objetivo específico. Estructurar la estrategia didáctica mediante el trabajo por proyectos.*

Las preguntas del proyecto Estadísticas de sexualidad se diseñan para abordar la construcción de gráficos estadísticos como ayuda visual que permita identificar la moda de una variable cualitativa nominal, la relación entre dos variables cualitativas, la representación de datos de una variable cuantitativa continua, la identificación de datos atípicos, la identificación de un intervalo modal para una variable cuantitativa continua, la representación de una variable categórica en términos porcentuales y la representación de las medidas resumen (media y mediana) con respecto a una variable cuantitativa discreta.

Se considera que estas preguntas utilizadas en el proyecto tienen interés desde la formación estadística porque se diseñan considerando un tema, un objetivo y una propuesta de solución teniendo en cuenta un conjunto de datos contextualizados (a partir del análisis a priori descrito en la sección 3.2.3).

- *Objetivo específico. Implementar y evaluar la estrategia didáctica.*

Al analizar los resultados de la estrategia didáctica, se observa que no todos los estudiantes deciden realizar gráficos estadísticos para la representación de los datos. Además, algunos estudiantes que construyen gráficos estadísticos suelen omitir algún elemento estructural (título, etiquetas, rótulos) o no tener en cuenta sus convenios de construcción, lo que puede generar interpretaciones incorrectas al momento de su lectura.

En el análisis de los gráficos producidos por los estudiantes en la prueba diagnóstica y en el proyecto Estadísticas de sexualidad (ver sección 4.1.3) se definen algunos niveles de respuesta con el fin de clasificar a cada uno de estos gráficos. Específicamente, con los resultados del proyecto se concluye que todos los estudiantes, que elaboran gráficos estadísticos, logran reconocer la información a representar para resolver las tres primeras preguntas. Sin embargo, se presentan dificultades enunciadas en investigaciones anteriores, como la realizada por Pedro Arteaga (2008), con respecto a la falta de interpretación de un gráfico o a su interpretación errónea por no reconocer sus elementos estructurales o no apreciar el impacto que tienen estos elementos en la presentación de la información.

En particular, el total de estudiantes que elaboran un gráfico para identificar los posibles valores atípicos reconocen la necesidad y pertinencia de utilizar un boxplot. Sin embargo,

teniendo en cuenta la complejidad del gráfico y la cantidad de conceptos preliminares asociados a su elaboración, la mayoría de ellos presentan algunos errores en los cálculos o en algunos convenios de su construcción.

Por otro lado, al analizar las respuestas de la prueba diagnóstica, antes y después de la intervención con el Trabajo por Proyectos, se observa que, en general, la intervención tuvo un efecto positivo en los estudiantes. Sólo en una de las 7 preguntas analizadas (pregunta 4), la probabilidad estimada de pasar de respuesta correcta a incorrecta fue mayor que la probabilidad estimada de pasar de respuesta incorrecta a correcta $0.5 = 4/8$ vs $0.4 = 9/15$. Es posible que en esta pregunta los estudiantes no utilicen correctamente las operaciones para el cálculo de porcentajes o no reconozcan la relación que existe entre las frecuencias relativas de una tabla de frecuencias y sus respectivas cantidades porcentuales (por el contrario, es posible que interpreten la frecuencia absoluta como el equivalente a la cantidad porcentual).

Por último, cabe resaltar que la resolución del proyecto *Estadísticas de sexualidad* permite realizar una lectura, interpretación y análisis de datos estadísticos, mostrando la utilidad de la estadística descriptiva en el análisis de situaciones cotidianas y la pertinencia de utilizar un análisis estadístico según el tipo de variable.

5.2. Recomendaciones

Con la implementación y evaluación de esta estrategia didáctica se sugieren las siguientes recomendaciones para aquellos educadores o investigadores que deseen poner en práctica la propuesta de este trabajo:

- Con el análisis realizado de las respuestas correspondientes a la cuarta pregunta de la prueba diagnóstica, antes y después de la intervención del trabajo con proyectos, se puede interpretar (de manera descriptiva) que hubo un efecto negativo en los estudiantes porque la probabilidad estimada de pasar de respuesta correcta a incorrecta fue mayor que la probabilidad estimada de pasar de respuesta incorrecta a correcta. Desde el punto de vista didáctico se mencionan algunas posibles causas para este resultado como presentar errores en los cálculos de porcentajes (por parte de los estudiantes). Desde el punto de vista estadístico, como se hizo el análisis pregunta por pregunta, se pierden algunas conexiones entre ellas. Es posible que un análisis, por ejemplo, de correspondencias múltiples, aunque descriptivo, nos ayude a explicar la

conexión de las respuestas, a la pregunta 4, de estos cuatro estudiantes, con otras respuestas. Este análisis queda como trabajo futuro.

- Para la lectura y análisis de un conjunto de datos se puede utilizar un software estadístico como Excel o R-Project, pero se recomienda utilizar R-Project porque se ajusta a los contenidos disciplinares propuestos en este trabajo teniendo en cuenta que se encontraron algunos errores técnicos en Excel como graficar un histograma utilizando barras no adosadas.
- Aunque el software R-Project utiliza un lenguaje de programación que puede resultar poco amigable a los estudiantes durante la primera impresión, se recomienda realizar varias actividades introductorias y de exploración con respecto a su entorno de trabajo y sus funciones (como el ingreso de datos, la lectura de un archivo de datos o la elaboración de gráficos estadísticos y sus respectivas configuraciones de color, tamaño de escala, asignación de rótulos, entre otros) para que los comandos cobren sentido en la interpretación y análisis de datos.
- En la pregunta 5 de la prueba diagnóstica, se agradece la observación de la docente evaluadora Emilse Gómez Torres con respecto a que la tabla propuesta difiere de las presentadas en el marco disciplinar. Para correcciones futuras, se recomienda utilizar sólo sus frecuencias absolutas.

A. Anexo: Prueba diagnóstica

En esta prueba encontrarás 10 preguntas referentes a estadística. Las primeras siete preguntas están conformadas por un enunciado y cuatro opciones de respuesta (A, B, C, D). Entre las opciones presentadas existe solo una que satisface las exigencias del enunciado. Debes seleccionar la respuesta correcta y marcarla en tu hoja de respuestas rellenando el óvalo correspondiente a la letra que identifica la opción elegida.

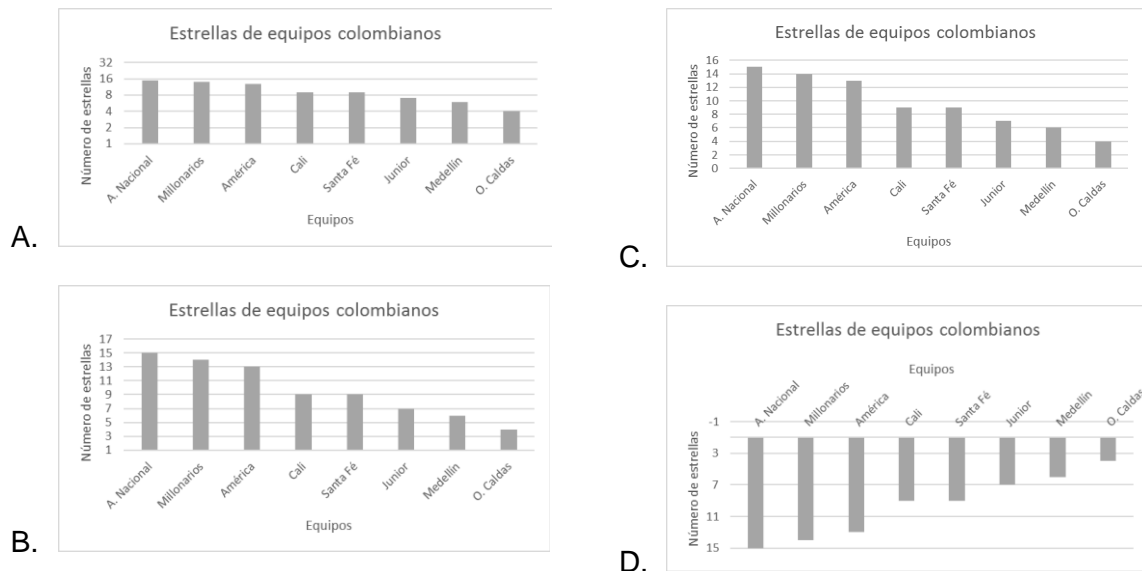
Las últimas tres preguntas deben ser contestadas en la hoja entregada con sus respectivos procedimientos. Es necesario tener los utensilios regla y transportador, pero no se permitirá el uso de calculadoras o celulares.

Responda la pregunta 1 con la siguiente información

En la siguiente tabla encuentras información sobre el número de estrellas que han obtenido equipos de fútbol en el campeonato nacional:

Estrellas	15	14	13	9	9	7	6	4
Equipo	A. Nacional	Millonarios	América	Cali	Santa Fe	Junior	Medellín	O. Caldas

1. Cuál de los siguientes gráficos representa correctamente la información de la tabla:



Responda las preguntas 2, 3 y 4 con la siguiente información

Se encuestó a un grupo de estudiantes, hombres y mujeres, por su música preferida. La información se resumió en la siguiente tabla:

		Música Preferida				
		Rock	Salsa	Rap	Reggaetón	Vallenato
Sexo Biológico	Hombres	14	4	12	6	2
	Mujeres	3	9	5	11	7

2. ¿Cuántos estudiantes, en total, fueron encuestados?
- A. 38
B. 35
C. 73
D. 17
3. Si los géneros *Rock* y *Rap* representan a Norteamérica, la *Salsa* y el *Reggaetón* a Centroamérica y el *Vallenato* a Colombia, ¿cuál región tiene mayor representación de estudiantes?
- A. Norteamérica con 26 estudiantes
B. Norteamérica con 34 estudiantes
C. Centroamérica con 30 estudiantes
D. Centroamérica con 39 estudiantes
4. El profesor solicitó a sus estudiantes usar la información de la tabla de frecuencias para asignar porcentajes a cada uno de los géneros musicales. La tabla que presenta correctamente esta información es:

A.

Género	Frecuencia	Porcentaje
Rock	17	23.28%
Salsa	13	17.80%
Rap	17	23.28%
Reggaetón	17	23.28%
Vallenato	9	12.32%

C.

Género	Frecuencia	Porcentaje
Rock	17	4.29%
Salsa	13	5.61%
Rap	17	4.29%
Reggaetón	17	4.29%
Vallenato	9	8.11%

B.

Género	Frecuencia	Porcentaje
Rock	17	17%
Salsa	13	13%
Rap	17	17%
Reggaetón	17	17%
Vallenato	9	9%

D.

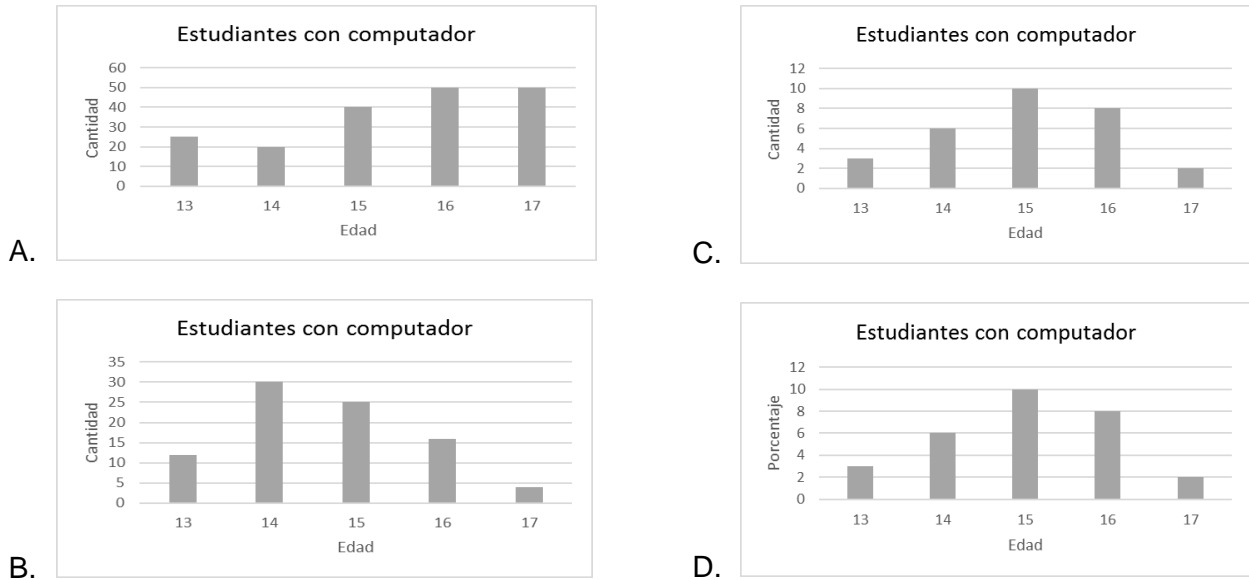
Género	Frecuencia	Porcentaje
Rock	14	36.8%
Salsa	4	10.5%
Rap	12	31.5%
Reggaetón	6	15.7%
Vallenato	2	5.26%

Responda la pregunta 5 con la siguiente información

En la siguiente tabla se muestra información acerca de los estudiantes del ciclo IV que tienen computador en su casa.

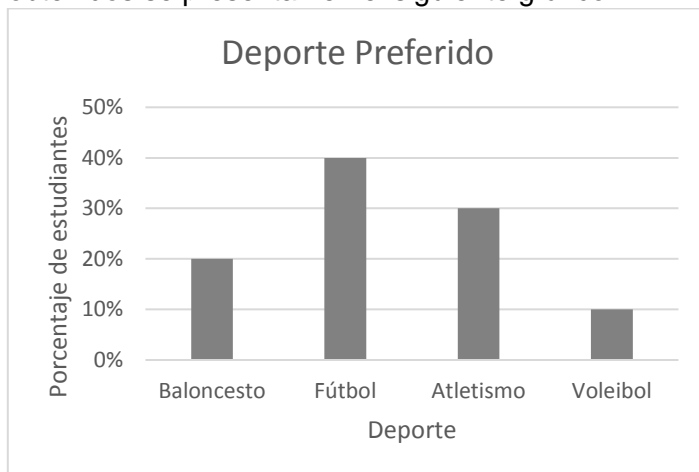
Edad	Porcentaje de estudiantes que tienen computador por edad	Número de estudiantes que tienen computador	Total de estudiantes
13	25%	3	12
14	20%	6	30
15	40%	10	25
16	50%	8	16
17	50%	2	4

5. La gráfica que ilustra la cantidad de estudiantes que tienen computador es:



Responda las preguntas 6 y 7 con la siguiente información

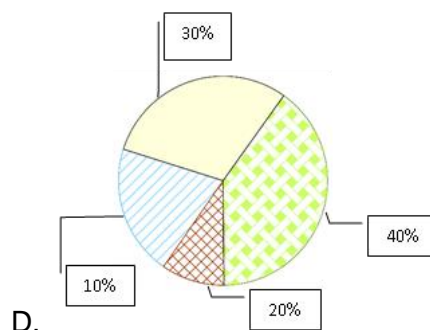
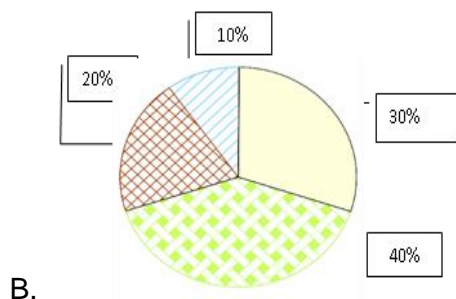
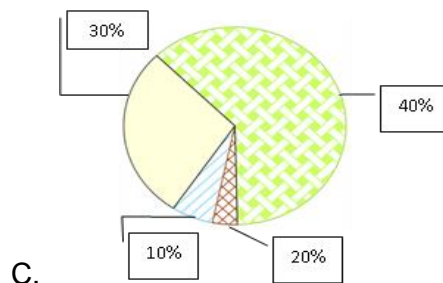
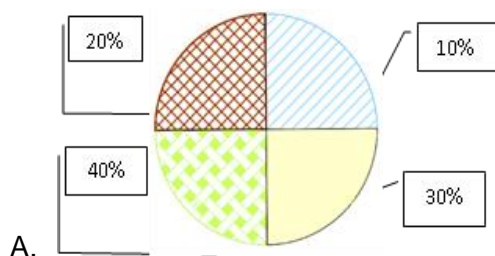
Se encuestó a un grupo de estudiantes por el deporte preferido. Los resultados obtenidos se presentan en el siguiente gráfico:



6. Usando la información del gráfico anterior se quiere construir un gráfico circular (también conocido como gráfico de pastel o torta) utilizando las siguientes convenciones:



El gráfico correcto es:



7. Si hay un total de 180 estudiantes, es correcto afirmar que:
- A. 20 estudiantes prefieren el Baloncesto.
 - B. 40 estudiantes prefieren el Atletismo y el Voleibol.
 - C. 72 estudiantes prefieren el Fútbol.
 - D. 6 estudiantes prefieren el Atletismo.

Responda la pregunta 8 con la siguiente información

8. En la tabla que se presenta a continuación encuentra información sobre los hábitos de lectura de noticias de un grupo de adolescentes y adultos. Complete los espacios vacíos y construya un gráfico circular (también conocido como torta, pastel o de sectores) para la variable *Lee noticias*.

	Lee noticias	No lee noticias	Total
Adolescentes	230	970	
Adultos	850		1400
Total			2600

Responda las preguntas 9 y 10 con la siguiente información

En la tabla que se presenta a continuación encuentra información sobre un grupo de pacientes con una enfermedad y los efectos del tratamiento con dos tipos de medicamentos (Medicamento A y Medicamento B).

	Curados	No curados	Total
Medicamento A	95	85	180
Medicamento B	90	70	160
Total	185	155	340

9. Construya un gráfico de barras para la variable Medicamento con la información de los pacientes de este grupo curados.
10. Construya un gráfico de barras (en términos porcentuales) para la variable Enfermedad (curado, no curado) respecto a los tratados con el medicamento A.

B. Anexo: Hoja de códigos. Estadísticas de sexualidad

Variable	Descripción	Tipo de variable	Rango	Categorías
Código	Código de la encuestada	Cualitativa	n.a.	
Edad	Edad en años de la encuestada	Cuantitativa	25-30	
N_Edu	Nivel educativo más alto	Cualitativa	n.a.	0: Sin educación 1: Primaria 2: Secundaria 3: Mayor
A_Acad	Educación en años individuales	Cuantitativa	0-19	
N_Miembros	Número de miembros del hogar	Cuantitativa	1-11	
S_Hogar	Sexo biológico de la cabeza de hogar	Cualitativa	n.a.	1: Hombre 2: Mujer
Estrato	Estrato socio económico	Cualitativa	n.a.	
T_Niños	Total de hijos nacidos	Cuantitativa	0-5	
C_Ovu	Conocimiento del periodo fértil en el ciclo ovulatorio	Cualitativa	n.a.	1: Durante la menstruación 2: Después de finalizar la menstruación 3: En medio del ciclo 4: Antes de que comience la menstruación 5: En cualquier momento 6: Otro 8: No sabe
M_Ant	Método anticonceptivo actual	Cualitativa	n.a.	0: No usa 1: Píldora 2: DIU 3: Inyección 4: Diafragma 5: Condón 6: Esterilización femenina 7: Esterilización masculina 8: Periodo de abstinencia

				9: Retiro (coito interrumpido) 10: Otro 11: Implantes 12: Abstinencia 13: Lactancia y amenorrea 14: Condón femenino 15: Espermicidas
Peso	Peso del entrevistado (Kilos)	Cuantitativa	36-82.5	
Altura	Altura del entrevistado (Centímetros)	Cuantitativa	140.8-170.6	
E_PR	Edad del primer coito	Cuantitativa	12-23	0: No tiene relaciones sexuales 96: En la primera unión 97: Inconsistente 98: No lo sabe
N_IHijos	Número ideal de hijos	Cuantitativa	0-5	98: No lo sabe
E_TS	Ha escuchado de enfermedades de transmisión sexual	Cualitativa	n.a.	0: No 1: Si 8: No lo sabe
H_Homosexual	¿Qué haría si se descubre que uno de sus hijos es homosexual? Nota: la palabra hijos hace referencia tanto a hombres como a mujeres.	Cualitativa	n.a.	1: Lo retiraría de la casa 2: Retiraría ayuda económica 3: Lo envió a ver un psicólogo 4: Lo envió a ver un doctor 5: Apoyaría 6: Otro 7: Aceptaría 8: Proporcionaría consejo/diálogo

n.a.=no aplica

C. Anexo: Proyecto. Estadísticas de sexualidad

Se desea realizar un estudio acerca de temas de sexualidad en jóvenes utilizando información de la Encuesta Nacional de Demografía y Salud. En el año 2010, durante un estudio realizado acerca de la salud sexual y reproductiva se entrevistaron a 53521 mujeres en edad fértil. No obstante, para este ejercicio académico, se utiliza una sub-base con sólo 80 registros, con edades entre los 25 y 30 años, y con residencia en Bogotá.

Con los datos entregados, responda las siguientes preguntas utilizando los cálculos, gráficos o resúmenes estadísticos que considere convenientes:

1. ¿Cuál es el método anticonceptivo típico utilizado por las mujeres? ¿Cómo llega a esa respuesta?
2. ¿Existe alguna relación entre el sexo biológico de la cabeza de hogar con lo que se puede hacer si se descubre que uno de los hijos es homosexual? ¿Se puede representar la anterior información?

Nota del docente: La palabra hijos, en este contexto, hace referencia tanto a hombres como a mujeres. Esto es ajeno al docente.

3. ¿Existen valores atípicos en la altura de las mujeres entrevistadas? Si los hay, ¿son valores atípicos pequeños o grandes?
4. ¿Cómo se puede representar la altura de las mujeres entrevistadas?
5. ¿Cuál es el intervalo del peso más común entre las mujeres entrevistadas?
6. ¿Se puede representar en un gráfico la información referente al conocimiento del periodo fértil en el ciclo ovulatorio utilizando términos porcentuales? Y si ahora se restringe la información entre las mujeres con nivel educativo secundaria, ¿se puede realizar el gráfico?
7. Calcule las medidas resumen del número hijos con respecto a cada nivel educativo, usando dos medidas resumen diferentes: la media y la mediana. ¿Qué diferencias observa entre las dos medidas resumen? ¿Cuál medida elegir, entre la media o la mediana, para representar la medida resumen del número hijos, según su nivel educativo?

Nota: resuelva el proyecto y realice un informe adjuntando los cálculos y gráficos que considere pertinentes.

D. Anexo: Actividad de reconocimiento en R-Project

En este anexo se presentan algunos comandos básicos del software estadístico R-Project usados en la estadística descriptiva (ver tabla 32). Posteriormente, se describe una actividad de reconocimiento del software a partir del ingreso de datos, el cálculo de resúmenes estadísticos y la elaboración de algunos gráficos.

Tabla 32: Algunos comandos de R-Project utilizados en estadística descriptiva

Comando	Función
<i>attach(X)</i>	Referencia los nombres de los encabezados en las columnas del <code>data.frame</code> .
<i>sapply(X, class)</i>	Indica la clase de variable estadística para cada objeto (columna).
<i>str(X)</i>	Muestra la estructura de cada objeto (Variable).
<i>mean(X)</i>	Media aritmética o promedio de la variable cuantitativa X .
<i>median(X)</i>	Mediana de la variable cuantitativa X .
<i>summary(Y)</i>	Estadísticos resumen de la variable Y .
<i>by(X, Y, summary)</i>	Estadísticos resumen de la variable Y (continua), con respecto a cada una de las categorías de la variable X .
<i>sd(Y)</i>	Desviación estándar de la variable cuantitativa Y (<u>S</u> tandard <u>D</u> eviation).
<i>var(X)</i>	Varianza de la variable cuantitativa X .
<i>plot(X, Y)</i>	Gráfico de dispersión Y vs X , con la variable Y en el eje vertical y la variable X en el eje horizontal.
<i>barplot(X)</i>	Gráfico de barras de la variable categórica X .
<i>pie(X)</i>	Gráfico de pastel (también conocido como de sectores o circular) de la variable categórica X .
<i>table(X)</i>	Tabla con frecuencias absolutas de la variable X .
<i>table(X, Y)</i>	Tabla de contingencia para las variables categóricas X y Y .
<i>hist(X)</i>	histograma de la variable continua X (número de intervalos dado por R).
<i>quantile(X)</i>	Cuantiles de la variable cuantitativa X .
<i>quantile(X,0.25)</i>	Cálculo del primer cuartil de la variable cuantitativa X .
<i>quantile(X,0.75)</i>	Cálculo del tercer cuartil de la variable cuantitativa X .
<i>boxplot(Y)</i>	Gráfico de cajas de la Variable cuantitativa Y .
<i>nclass.Sturges(X)</i>	Número de intervalos por el método Sturges para la variable continua X .
<i>range(X)</i>	Muestra el valor mínimo y el máximo de la variable cuantitativa X .
<i>pretty(X,n)</i>	Divide la variable cuantitativa X en aproximadamente n intervalos iguales y muestra los puntos de corte para los intervalos de clase.

Actividad de reconocimiento

Para comenzar, se analizarán las notas de matemáticas de un grado séptimo (utilizando el software R-Project) mostradas en la siguiente tabla:

45	30	41	60	25
57	32	44	60	66
47	86	47	89	61
67	75	63	95	53
67	55	60	50	64

Instrucciones.

1. Se abre el programa R-Project.
2. Diríjase a **Archivo**. De click en **Nuevo Script** (Se abre una pequeña ventana).
 - i. Diríjase a **Ventanas** y dar click en **Divida verticalmente**.

Observación: Se quedan abiertas dos ventanas, una para el Script (Editor R) y otra para la Consola.

3. De click sobre la consola y presione Ctrl+L, para limpiar o borrar los datos de la ventana.
4. Luego, en la ventana del Script, se ingresan los datos utilizando el siguiente comando:

Notas=c(45,30,41,...,64)

La palabra “Notas” representa el nombre de la lista de datos y se debe tener en cuenta para todos los comandos que se utilicen en adelante.

El comando *c()* representa el ingreso de una lista de datos.

5. Presione la tecla F5 para ejecutar el comando en la consola. Este paso se repite cada vez que escriba una instrucción sobre el Script. Los resultados del comando aparecen en la ventana de la consola.

6. Después, ingrese el siguiente comando en el Script:

summary(Notas)

No olvide presionar F5. ¿Qué ha sucedido? ¿Comprende los resultados que aparecen en la consola?

7. ¿Cuál es el rango de las Notas? Recuerde que el rango corresponde a $x_{m\acute{a}x} - x_{m\acute{i}n}$. Ingrese el siguiente comando para conocer el valor máximo y el valor mínimo:

range(Notas)

8. Para conocer si alguno de los datos se repite, le pediremos a R que elabore una tabla de frecuencias absolutas con el siguiente comando:

table(Notas)

9. Adicionalmente, se pide que represente los datos por medio de un gráfico de barras utilizando el siguiente comando:

```
barplot(Notas)
```

Como puede observar, es un gráfico simple que le faltan etiquetas.

10. Para colocar las etiquetas al gráfico se utiliza los comandos *main=* para el título, *xlab=* para la etiqueta del eje x, *ylab=* para la etiqueta del eje Y, y *col=* para indicar el color de las barras, tal como se muestra en el siguiente comando:

```
barplot(Notas, col="Yellow",main="Notas", xlab="Puntaje", ylab="Frecuencia")
```

Recuerde que puede personalizar su gráfico con los colores que desee.

11. A continuación, represente los datos utilizando un gráfico de sectores (circular) con el siguiente comando:

```
pie(Notas)
```

También se puede añadir un título con:

```
pie(Notas,main="Notas")
```

Como puede notar en las anteriores instrucciones, con la “alta” variación de las notas no es conveniente utilizar el gráfico de barras o el gráfico circular. Por tanto, es mejor elaborar un gráfico que agrupe datos (hacer el tratamiento de una variable cuantitativa) como el histograma.

12. Ingrese el siguiente comando:

```
hist(Notas)
```

¿Cómo le puede cambiar el título?

13. Por último, realice un gráfico de caja y bigotes (boxplot) utilizando el siguiente comando:

```
boxplot(Notas)
```

¿Hay datos atípicos? ¿Cómo asignar nombre y etiquetas?

Con el comando *boxplot(Notas, outcol=2)*, se resalta el dato atípico con color rojo.

E. Anexo: Importar datos a R-Project

Cuando se trabaja con una base de datos puede resultar tedioso pasar dato por dato al programa R. Por tanto, como algunas de las bases se suelen trabajar en Microsoft Excel, hay unos ciertos comandos que nos permiten importar los datos directamente al programa R-Project.

Nota: en este ejercicio se toma como referencia una base de datos de los admitidos al programa de Física de la Universidad Nacional en el periodo 2016-1.

Instrucciones preliminares:

- i. Si la base de datos tiene formato xls (Excel), abra el archivo.
- ii. Diríjase a **Archivo** y de click en **guardar como**.
- iii. Diríjase a la pestaña donde se encuentran los diferentes formatos y de click en Texto (delimitado por tabulaciones)(*txt). Guarde el documento en Escritorio (preferiblemente) y recuerde su nombre (admitidos).

Instrucciones de importación a R:

1. Abra R-Project.
2. Diríjase a **Archivo** y abra **Nuevo Script**.
 - i. Diríjase a **Ventanas** y de click en **Divida verticalmente**.
3. De click sobre la consola y presione Ctrl+L, para limpiar o borrar los datos de la ventana.
4. Sobre la ventana de Consola, Diríjase a **Archivo** y de click en **Cambiar dir...**
 - i. Se abre una nueva ventana donde debe buscar la carpeta del archivo *txt. Que había guardado previamente en Excel. Por lo general, se debe dar click en Disco local (C), luego en Usuarios, luego en el administrador del equipo y, por último, en Escritorio. Finalmente presiona aceptar.
5. Luego, en la ventana del Script, se ingresan los siguientes comandos:

Comando	Función
<code>rm(list=ls())</code>	Para borrar datos anteriores.
<code>admitidos=read.table("admitidos.txt", head=TRUE, dec=",")</code>	Leer el archivo, donde <code>read.table</code> significa leer la table y <code>head=TRUE</code> significa reconocer los encabezados.
<code>head(admitidos)</code>	Muestra los primeros seis datos de la base.

<code>attach(admitidos)</code>	Referencia los nombres de los encabezados
<code>sapply(admitidos, class)</code>	Indica la clase de variable estadística para cada columna.
<code>str(admitidos)</code>	Muestra la estructura de cada variable.
<code>summary(admitidos)</code>	Muestra el resumen estadístico para la base de datos <i>admitidos</i> .

Funciones adicionales:

6. Si desea un histograma para una variable específica (por ejemplo: *hstMatematicas*), ingrese el siguiente comando:

```
hist(hstMatematicas)
```

7. Si desea una tabla de contingencia entre dos variables categóricas como *Género* y *Estrato*, ingrese el siguiente comando:

```
table(Genero,Estrato)
```

8. Además, si quiere representar la anterior información (dos variables), ingresa el siguiente comando:

```
barplot(table(Genero,Estrato),beside=TRUE)
```

9. Por último, cuando se quiere conocer el resumen estadístico de una variable con respecto a una variable categórica (teniendo en cuenta cada una de sus categorías), se escribe el siguiente comando ejemplo:

```
by(hstMatematicas, Estrato, summary)
```

R presenta el resumen estadístico del puntaje en matemáticas con respecto al estrato socioeconómico.

F. Anexo: Modificaciones a la prueba diagnóstica

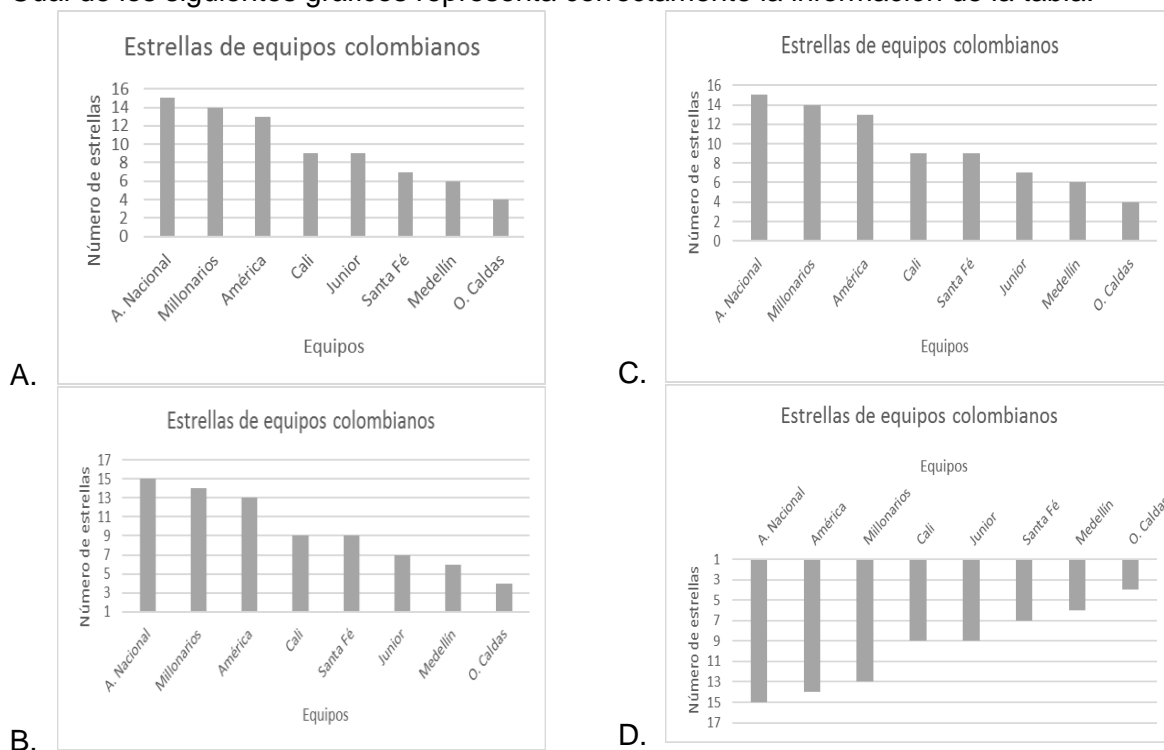
A continuación, se presentan sólo los puntos de la prueba diagnóstica que fueron modificados para el segundo momento de aplicación.

Responda la pregunta 1 con la siguiente información

En la siguiente tabla encuentras información sobre el número de estrellas que han obtenido equipos de fútbol en el campeonato nacional:

Estrellas	15	14	13	9	9	7	6	4
Equipo	A. Nacional	Millonarios	América	Cali	Santa Fe	Junior	Medellín	O. Caldas

1. Cuál de los siguientes gráficos representa correctamente la información de la tabla:



Responda las preguntas 2, 3 y 4 con la siguiente información

Se encuestó a un grupo de estudiantes, hombres y mujeres, por su música preferida. La información se resumió en la siguiente tabla:

		Música Preferida				
		Rock	Salsa	Rap	Reggaetón	Vallenato
Sexo Biológico	Hombres	14	4	12	6	2
	Mujeres	3	9	5	11	7

4. El profesor solicitó a sus estudiantes usar toda la información de la tabla de frecuencias para asignar porcentajes a la música preferida. La tabla que presenta correctamente esta información es:

A.

Género	Frecuencia	Porcentaje
Rock	17	23.28%
Salsa	13	17.80%
Rap	17	23.28%
Reggaetón	17	23.28%
Vallenato	9	12.32%

C.

Género	Frecuencia	Porcentaje
Rock	17	4.29%
Salsa	13	5.61%
Rap	17	4.29%
Reggaetón	17	4.29%
Vallenato	9	8.11%

B.

Género	Frecuencia	Porcentaje
Rock	17	17%
Salsa	13	13%
Rap	17	17%
Reggaetón	17	17%
Vallenato	9	9%

D.

Género	Frecuencia	Porcentaje
Rock	14	36.8%
Salsa	4	10.5%
Rap	12	31.5%
Reggaetón	6	15.7%
Vallenato	2	5.26%

Responda la pregunta 8 con la siguiente información

8. En la tabla que se presenta a continuación encuentra información sobre los hábitos de lectura de noticias de un grupo de adolescentes y adultos.
- Complete los espacios vacíos.
 - Construya un gráfico circular (también conocido como torta, pastel o de sectores) para la variable *Lee noticias*, en los dos grupos de edades.

	Lee noticias	No lee noticias	Total
Adolescentes	230	970	
Adultos	850		1400
Total			2600

Responda las preguntas 9 y 10 con la siguiente información

En la tabla que se presenta a continuación encuentra información sobre un grupo de pacientes con una enfermedad y los efectos del tratamiento con dos tipos de medicamentos (Medicamento A y Medicamento B).

	Curados	No curados	Total
Medicamento A	95	85	180
Medicamento B	90	70	160
Total	185	155	340

9. Construya un gráfico de barras para la variable "Medicamento" con la información de los pacientes del grupo "Curados".

Apéndice

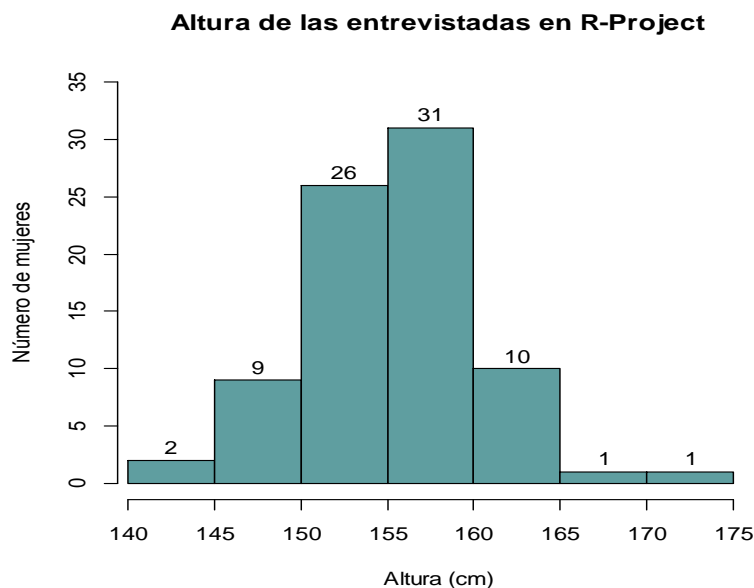
En el capítulo 3 se presentaron las preguntas del proyecto Estadísticas de sexualidad y una propuesta de solución para cada una de ellas. Específicamente, en el caso de las preguntas 3, 4 y 5 se trabaja dos variables de naturaleza cuantitativa continua por medio de un estudio para datos agrupados.

Como se dio la posibilidad de trabajar el proyecto utilizando los cálculos desde el software R-Project, en este apéndice se describe la solución que ofrece este software para los puntos 3, 4 y 5 dado que en sus cálculos utiliza otro método. Cabe mencionar que una de las desventajas encontradas es no poder visualizar el mínimo y el máximo en la tabla de distribución de frecuencias (teniendo en cuenta que el límite inferior de la primera clase es menor que el valor mínimo de los datos y el límite superior de la última clase es mayor que el máximo) y en sus respectivos histogramas. Además, la notación utilizada para los intervalos de clase es diferente a los utilizados en este trabajo dado que son abiertos por la izquierda y cerrados por la derecha.

3. ¿Cómo representar la altura de las mujeres entrevistadas?

Realizando el análisis de los datos utilizando el algoritmo del software R-Project, se obtiene el siguiente histograma utilizando el comando `hist(Altura)`:

Figura 24: Histograma para las alturas en R



R-Project utiliza en sus cálculos el comando $pretty(Altura, n)$ para dividir la *Altura* en aproximadamente n intervalos iguales, teniendo en cuenta que los puntos de división son valores redondeados. En este caso:

- El número de intervalos es $k = 7$.
- La amplitud es $a = 5$.

Tabla 33: Distribución de frecuencias para las alturas en R

<i>Alturas</i>	f_i	h_i	F_i	H_i
(140,145]	2	0.025	2	0.025
(145,150]	9	0.112	11	0.137
(150,155]	26	0.325	37	0.462
(155,160]	31	0.387	68	0.850
(160,165]	10	0.125	78	0.975
(165,170]	1	0.012	79	0.987
(170,175]	1	0.012	80	1.000

4. ¿Existen valores atípicos en la altura de las mujeres entrevistadas? Si los hay, ¿son valores atípicos pequeños o grandes?

Este software realiza los cálculos para datos enumerados, de manera que:

- Primer cuartil: con el comando $quantile(Altura, 0.25)$ R ordena ascendentemente los datos e identifica el índice (i -ésima estadística de orden) que corresponde al primer cuartil por medio de la siguiente expresión -por defecto-:

$$i = \frac{(n-1)p}{100} + 1 = \frac{(80-1)25}{100} + 1 = 20.75, \text{ donde } p \text{ corresponde a } 25 \text{ (número del percentil)}$$

y n es el total de datos.

El dato 20 corresponde a 151.6 y el dato 21 a 151.7. Luego, R-Project realiza la interpolación lineal para determinar el primer cuartil:

$$q_1 = 151.6 + \left(\frac{20.75-20}{21-20} \right) (151.7 - 151.6) = 151.675.$$

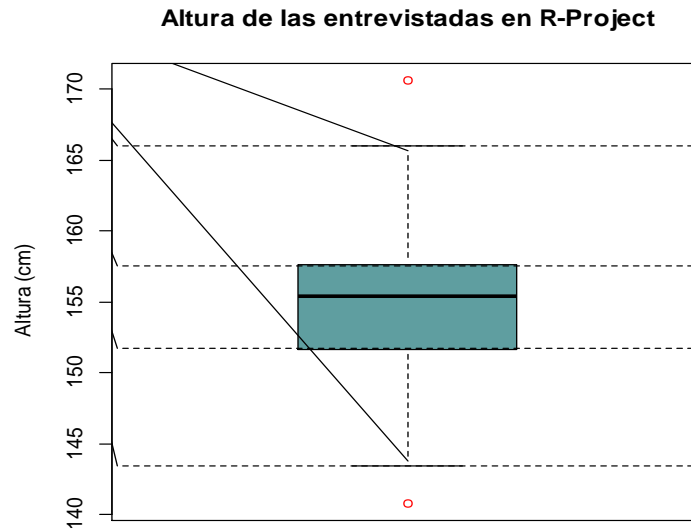
De manera análoga, para el segundo y tercer cuartil se utilizan los comandos $quantile(Altura, 0.50)$ (o también $median(Altura)$) y $quantile(Altura, 0.75)$, donde sus valores son $q_2 = 155.35$ y $q_3 = 157.55$, respectivamente.

- El rango intercuartílico corresponde a: $RI_{Altura} = q_3 - q_1 = 157.55 - 151.675 = 5.875$

- El bigote inferior se localiza utilizando la expresión $bigote.inf = \max\{x_{\min}, LI\}$, donde $LI = q_1 - (1.5)RI_{Altura} \approx 142.86$, de manera que $bigote.inf = 143.4$.
- Para el bigote superior se utiliza la expresión $bigote.sup = \min\{x_{\max}, LS\}$, donde $LS = q_3 + (1.5)RI_{Altura} \approx 166.36$, por tanto $bigote.sup = 166$.

Finalmente, el software construye el siguiente boxplot:

Figura 25: Boxplot para las alturas en R

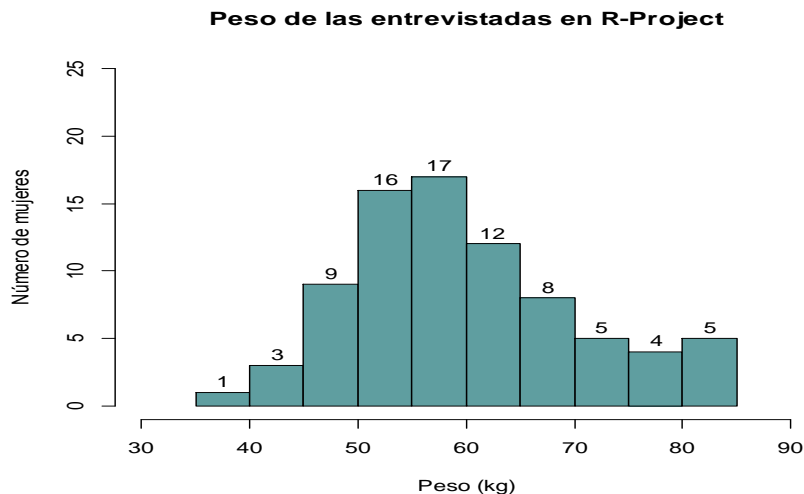


Con este gráfico, se pueden identificar dos valores atípicos: un valor atípico grande (170.6) y un valor atípico pequeño (140.8).

5. ¿Cuál es el intervalo del peso más común entre las mujeres entrevistadas?

Con el comando `hist(Peso)`, el software R construye el siguiente histograma de los datos:

Figura 26: Histograma para los pesos en R



Luego, con el comando $pretty(Peso, n)$, R-Project divide el *Peso* en 10 intervalos iguales, con amplitud $a = 5$.

Tabla 34: Distribución de frecuencias para los pesos en R

<i>Peso</i>	f_i	h_i	F_i	H_i	x_i
(35,40]	1	0.012	1	0.012	37.5
(40,45]	3	0.037	4	0.050	42.5
(45,50]	9	0.112	13	0.162	47.5
(50,55]	16	0.200	29	0.362	52.5
(55,60]	17	0.212	46	0.575	57.5
(60,65]	12	0.150	58	0.725	62.5
(65,70]	8	0.100	66	0.825	67.5
(70,75]	5	0.062	71	0.887	72.5
(75,80]	4	0.050	75	0.937	77.5
(80,85]	5	0.062	80	1.000	82.5

Con la tabla 34, se identifica el intervalo modal (55,60] y su respectiva moda (utilizando el cálculo para la marca de clase):

$$x_i = \frac{l_{i-1} + l_i}{2} = \frac{55 + 60}{2} = 57.5.$$

Bibliografía

- Aguilera, A. M. (2005). Análisis de tablas de contingencia bidimensionales. Madrid.
- Arteaga, P. (2008). Análisis de gráficos estadísticos elaborados en un proyecto de análisis de datos. Trabajo fin de Master. Departamento de Didáctica de la Matemática.
- Arteaga, P., Batanero, C., Cañadas, G., & Contreras, M. (2011). Las Tablas y Gráficos Estadísticos como Objetos Culturales. *Números. Revista de didáctica de las matemáticas*, 76, 55-67.
- Batanero, C., Arteaga, P. & Ruíz, B. (2009). Análisis de la complejidad semiótica de los gráficos producidos por futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas. *Enseñanza de las ciencias: revista de investigación y experiencias didácticas*, 28(1), 141-154.
- Batanero, C. & Díaz, C. (2011). Estadística con proyectos. Departamento de Didáctica de la Matemática. Granada.
- Bausela, E. (2004). La docencia a través de la Investigación-acción. *Revista Iberoamericana de Educación*, 35(1), 1-9. Recuperado de http://www.une.edu.ve/uneweb2005/servicio_comunitario/investigacion-accion.pdf.
- Coronado, J. (2007). Escalas de medición. *Paradigmas. Una revista disciplinar de investigación*, 2(2), 104 -125.
- Friel, S., Curcio, F. & Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education*, 32(2), 124-158.
- Gómez, M. (2006). Introducción a la metodología de la investigación científica. Editorial Brujas, 32-38.
- Grajales, L. (2016). Estadística descriptiva univariada, bivariada y multivariada. Notas de clase. Manuscrito no publicado. Universidad Nacional de Colombia.
- Hoffman, J. I. (1976). The incorrect use of Chi-square analysis for paired data. *Clinical and experimental immunology*, 24(1), 227-229.

ICFES, (2013). Pruebas Saber 3°, 5° y 9°. Lineamientos para las aplicaciones muestral y censal 2013. Bogotá.

ICFES, (2015). Día e. Caja de materiales. Matriz de referencia. Recuperado de http://aprende.colombiaaprende.edu.co/ckfinder/userfiles/files/articles-52712_matriz_m.pdf

ICFES, (2016). Lineamientos generales para la presentación del examen de Estado Saber 11°. Edición 2.

INEI [Instituto Nacional de Estadística e Informática] (2009). Guía para la presentación de gráficos estadísticos. Lima.

MEN, (2006). *Estándares básicos de competencias en lenguaje, matemáticas, ciencias y ciudadanas*. Bogotá.

Ojeda, G., Ordoñez, M. & Ochoa, L. (2011). Encuesta Nacional de Demografía y Salud 2010. Bogotá: Profamilia.

Sosa, J., Ospina, L. & Berdugo, E. (2013). Estadística descriptiva y probabilidades. Bogotá: Universidad Externado de Colombia, 17-32.