

PROPUESTA PARA PROMOVER EL RAZONAMIENTO  
ESTADÍSTICO DE ESTUDIANTES DE GRADOS NOVENO Y  
UNDÉCIMO AL COMPARAR CONJUNTOS DE DATOS MEDIANTE  
GRÁFICOS DE CAJA

EDILMA CHAPARRO PAVA  
JOSÉ ALCIDES ROMERO MARTÍNEZ

UNIVERSIDAD PEDAGÓGICA NACIONAL DE COLOMBIA  
BOGOTÁ D.C.

2012

PROPUESTA PARA PROMOVER EL RAZONAMIENTO  
ESTADÍSTICO DE ESTUDIANTES DE GRADOS NOVENO Y  
UNDÉCIMO AL COMPARAR CONJUNTOS DE DATOS MEDIANTE  
GRÁFICOS DE CAJA

EDILMA CHAPARRO PAVA  
JOSÉ ALCIDES ROMERO MARTÍNEZ

Trabajo de grado para optar por el título de Especialista en Educación  
Matemática

ASESOR: FELIPE JORGE FERNÁNDEZ HERNÁNDEZ

UNIVERSIDAD PEDAGÓGICA NACIONAL DE COLOMBIA  
FACULTAD DE CIENCIA Y TECNOLOGÍA  
DEPARTAMENTO DE MATEMÁTICAS  
BOGOTÁ D.C.

2012

Nota de aceptación

---

---

---

Presidente del Jurado

---

Jurado

---

Jurado

Bogotá, 6 de Noviembre de 2012

Para todos los efectos, declaramos que el presente trabajo es original y de nuestra total autoría; en aquellos casos en los cuales hemos requerido del trabajo de otros autores o investigadores, hemos dado los respectivos créditos.

## AGRADECIMIENTOS

*A Dios por darme la oportunidad de alcanzar una más de mis metas, ser mi fortaleza en cada instante, permitirme compartir con un gran compañero y apoyo incondicional.*

*A mi madre, Margarita Pava que aún desde la distancia me brinda fuerzas y motivación para seguir adelante.*

*A toda mi familia, que son la más grande bendición que Dios me ha dado.*

*Al profesor Felipe Fernández por sus orientaciones y colaboración para la elaboración del proyecto: ¡Dios lo bendiga!*

*A José Alcides Romero Martínez por ser un excelente amigo y compañero de trabajo, compartir cada uno de los instantes ya sea de alegría por nuestros triunfos o las preocupaciones por los trabajos que teníamos que entregar. Dios lo siga bendiciendo, muchas gracias, trabajar con usted fue muy placentero, espero que la vida nos permita continuar y cumplir nuestra próxima meta.*

*A Omar Barrera Gutiérrez: Amor muchas gracias por escucharme y tenerme paciencia cada vez que por mi estrés quería llorar o escuchar una palabra de ánimo.*

*A mi amiga Angelita, que siempre estuvo dispuesta a colaborar en todo sentido sin importar la hora, o el cansancio que pudiera tener, haces alusión a tu nombre. Dios derrame muchas bendiciones sobre ti.*

*A mis estudiantes por formar parte fundamental de este proyecto.*

*A cada una de las personas que de una u otra forma aportaron para que hoy logre alcanzar esta meta, muchas gracias.*

*Edilma Chaparro Pava*

## AGRADECIMIENTOS

En este importante momento de nuestras vidas; agradezco al creador por la culminación de esta importante meta.


En este instante deseo dedicar este trabajo a mi querida sobrina quien es la luz de mis ojos; un angelito que me ha dado un motivo más para seguir emprendiendo el camino de mi vida.

Agradezco a mi padre por su incondicional apoyo a lo largo de mi vida; a mi madre por su gran amor y sus oraciones y a mis queridas hermanas por su gran afecto y cariño.

En el paso por la especialización estaré profundamente agradecido por el apoyo brindado por el profesor Felipe Fernández; quien con su calidad humana nos demuestra que el conocimiento no es la única condición para ser un gran maestro.


Finalmente a mi amiga Edilma Chaparro con quien compartimos los mejores momentos en esta etapa de mi vida; cuya amistad espero poder contar el resto de mi vida; a ella le deseo lo mejor.

José Alcides Romero Martínez

|  |   |  |
|--|---|--|
| <br>UNIVERSIDAD PEDAGÓGICA NACIONAL<br><small>Formación de Maestros</small> | <b>FORMATO</b>                              |  |
|  | <b>RESUMEN ANALÍTICO EN EDUCACIÓN - RAE</b> |  |
| Código: FOR020GIB  | Versión: 01                                 |  |
| Fecha de Aprobación: 10-10-2012  | Página 1 de 4                               |  |

| <b>1. Información General</b> |   |
|-------------------------------|---|
| <b>Tipo de documento</b>      | Trabajo de Grado de Especialización   |
| <b>Acceso al documento</b>    | Universidad Pedagógica Nacional. Biblioteca Central   |
| <b>Título del documento</b>   | Propuesta para promover el razonamiento estadístico de estudiantes de grados noveno y undécimo al comparar conjuntos de datos mediante gráficos de caja                 |
| <b>Autor(es)</b>              | CHAPARRO, Edilma; ROMERO, José Alcides  |
| <b>Director</b>               | FERNÁNDEZ HERNÁNDEZ, Jorge Felipe   |
| <b>Publicación</b>            | Bogotá. Universidad Pedagógica Nacional, 2012. 89 páginas   |
| <b>Unidad Patrocinante</b>    | Universidad Pedagógica Nacional, Facultad de Ciencia y Tecnología, Departamento de Matemáticas  |
| <b>Palabras Claves</b>        | Interpretación de gráficos estadísticos, Análisis Exploratorio de Datos, Razonamiento estadístico, gráfico de caja, niveles y elementos de razonamiento, Taxonomía SOLO |

| <b>2. Descripción</b>   |
|---|
| <p>El presente trabajo busca dar respuesta a la pregunta: ¿Cómo promover el desarrollo del razonamiento estadístico de los estudiantes de grado noveno y undécimo a través del análisis de distribuciones de datos representados por gráficos de caja? para ello, se abordó entre otras cosas el estudio de las siguientes temáticas:</p> <ul style="list-style-type: none"> <li>• Naturaleza y características educativas del Análisis Exploratorio de Datos (AED), especialmente la técnica gráfica conocida como gráfico de cajas.</li> <li>• Caracterización del Razonamiento Estadístico</li> <li>• Aplicación de la taxonomía SOLO como modelo jerárquico en educación estadística.</li> <li>• Categorización del razonamiento estadístico de los estudiantes empleando el modelo diseñado.</li> <li>• Pautas para diseñar actividades de aprendizaje aplicadas en el gráfico de caja.</li> <li>• Aplicación a los estudiantes el cuestionario evaluativo diseñado por Palacios y Zambrano (2009)</li> </ul> <p>Se elaboraron y aplicaron cuatro guías de aprendizaje para introducir las medidas de tendencia central, los cuartiles, construcción, lectura e interpretación de un gráfico de caja, cada una planeada para sesiones de dos horas aproximadamente. La recolección de los datos se fundamentó en la aplicación de un cuestionario evaluativo que consta de dos situaciones problema de las cuales se presentan 10 ítems en total, sobre una muestra de 35 estudiantes del grado noveno de la IED Técnico Comercial Mariano Ospina Rodríguez de Guasca y 56 estudiantes del grado once de la IED General Carlos Albán de Albán. Los resultados obtenidos de esta aplicación se describieron y organizaron con base en las categorías planteadas por Palacios y Zambrano (2009) que constaron de dos variables, por un lado siete elementos de razonamiento y por otro cuatro niveles de razonamiento fundamentados en la taxonomía SOLO</p> |

|  |   |
|--|---|
| <br>UNIVERSIDAD PEDAGÓGICA NACIONAL<br><small>Formación de Maestros</small> | <b>FORMATO</b>                              |
|  | <b>RESUMEN ANALÍTICO EN EDUCACIÓN - RAE</b> |
| Código: FOR020GIB  | Versión: 01                                 |
| Fecha de Aprobación: 10-10-2012  | Página 2 de 4                               |

propuesta por Biggs y Collis (1982); realizando algunos cambios en los niveles de razonamiento, pues se encontraron respuestas de los estudiantes que no se ajustaban a estos niveles.

### 3. Fuentes

Las fuentes consultadas más relevantes que sirvieron de base para el desarrollo del proyecto fueron:

Batanero, C; Estepa, A; Díaz, G. (1991). Análisis Exploratorio de Datos: sus posibilidades en la enseñanza secundaria. *Suma*, n° 9, 1991: 25-31.

Biggs, J y Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Cobo, B. (2003). Significado de las medidas de posición central para los estudiantes de Secundaria. Universidad de Granada. Departamento de Didáctica de la Matemática.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382-393.

Gómez, P. (2002). Análisis didáctico y diseño curricular en matemáticas. *Revista EMA*, 7(3), 251-292.

Jones, G. Langrall, C. Money, E. y Thornton, C. (2004). Models of development. En D. Ben- Zvi y J. Garfield (eds), *The challeng of thinking*, pp 97 -117. Dordrecht: kluwer Academic Publishers.

Medina, A. (1998). *Didáctica de las matemáticas: Estrategias Metodológicas para la Enseñanza y el Aprendizaje de las Matemáticas*. Duitama: Universidad Pedagógica y Tecnológica de Colombia.

Ministerio de Educación Nacional (2006). *Estándares Básicos de Competencias en Matemáticas*. Bogotá: Ministerio de Educación Nacional [MEN].

Tukey, J. (1977) *Exploratory Data Analysis*. Editorial Addison Wesley.

Velleman, P. y Hoaglin D. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press.

Zambrano, W. y Palacios, R. (2009). *Razonamiento estadístico de estudiantes de secundaria al comparar conjunto de datos representados mediante gráficos de caja* (Tesis de maestría). Bogotá: Universidad Pedagógica Nacional.

### 4. Contenidos


El presente trabajo se encuentra organizado en las siguientes secciones a saber:

**JUSTIFICACIÓN:** Hace referencia a la necesidad que se tiene del conocimiento estadístico en cualquier contexto; en la parte educativa se busca generar en los estudiantes un acercamiento al razonamiento estadístico atendiendo a las propuestas curriculares del Ministerio de Educación Nacional y basados en las sugerencias de Palacios y Zambrano (2009), que proponen al gráfico de caja como una herramienta esencial del análisis exploratorio de datos.

**DEFINICIÓN DEL PROBLEMA:** Se inicia referenciado los diferentes gráficos que se pueden utilizar en el análisis exploratorio de datos, resaltando el gráfico de caja como objeto de estudio, que permite visualizar fácilmente la información y como una herramienta que fortalece en los estudiantes el razonamiento estadístico. Así como, los objetivos de la investigación y la pregunta: ¿Cómo promover el desarrollo del razonamiento estadístico de los estudiantes de grado noveno y undécimo a través del análisis de distribuciones de datos representados por gráficos de caja? que direcciona este proyecto.

**MARCO CONCEPTUAL:** Este capítulo presenta los referentes conceptuales generales que son fundamento del presente trabajo, como lo es, el análisis exploratorio de datos, el gráfico de caja y sus diferentes tipos, elementos y niveles de la taxonomía SOLO como modelo de



|   |   |
|---|---|
| <br>UNIVERSIDAD PEDAGÓGICA<br>NACIONAL<br><small>Formación de Maestros</small> | <b>FORMATO</b>                              |
|   | <b>RESUMEN ANALÍTICO EN EDUCACIÓN - RAE</b> |
| <b>Código: FOR020GIB</b>  | <b>Versión: 01</b>                          |
| <b>Fecha de Aprobación: 10-10-2012</b>  | <b>Página 3 de 4</b>                        |

caracterización del razonamiento estadístico, además de la propuesta de instrucción planteada para la enseñanza de los gráficos de caja.

**ENFOQUE METODOLÓGICO:** En esta sección se muestra la metodología que direccionó el desarrollo de este trabajo.

**RESULTADOS:** A partir de la implementación de la propuesta de instrucción y del instrumento de evaluación se realizó un análisis de los resultados obtenidos por los estudiantes con el fin de determinar el nivel de razonamiento estadístico. Este análisis comienza presentando ejemplos de respuestas obtenidas de cada nivel por ítem del cuestionario, posteriormente se realiza un análisis comparativo entre los estudiantes de grados undécimo, noveno y del grupo analizado en el proyecto de Palacios y Zambrano (2009)

**CONCLUSIONES:** Este capítulo está conformado por cinco secciones: conclusiones respecto a los objetivos, a la propuesta de instrucción, a los niveles de categorización del razonamiento, al instrumento de evaluación y a los resultados.

Finalmente se presentan las referencias bibliográficas que fueron el sustento conceptual de esta propuesta y los anexos que respaldan los resultados del mismo.

### 5. Metodología

El proyecto se desarrolló bajo la siguiente secuencia de actividades organizadas en tres fases:

Fase I. Análisis preliminares y propuesta de instrucción

- Decisiones acerca del gráfico de caja
- Categorías de análisis planteadas por Palacios y Zambrano
- Propuesta de instrucción

Fase II. Implementación y recolección de información

- Muestra de estudio
- Cronograma de aplicación


Fase III. Análisis de resultados

- Revisión y ajuste de las categorías
- Categorías ajustadas (Validación de categorías)
- Sistematización y análisis de resultados
- Comparación de poblaciones

### 6. Conclusiones

El empleo de categorías de análisis, posibilita un proceso de evaluación más preciso y formativo evidenciando de manera metódica y organizada las fortalezas y debilidades del proceso de enseñanza. En particular, se logró la caracterización y análisis del desempeño de los estudiantes al comparar dos distribuciones de datos organizados en gráficos de caja a partir de la cual se deduce la necesidad de revisar las tareas relacionadas con la comparación simultánea de diferentes conjuntos de datos.

La evaluación del razonamiento y pensamiento estadístico se hizo con base en el cuestionario de evaluación diseñado por Palacios y Zambrano (2009). Las respuestas de los estudiantes a este cuestionario fueron el punto de partida para iniciar la tarea de caracterizar el razonamiento y pensamiento estadístico al abordar la clasificación de las respuestas obtenidas, de acuerdo a cada uno de los elementos de razonamiento propuestos en la categorización del modelo SOLO, en los niveles: Preestructural, Uniestructural, Multiestructural y Relacional. Sin embargo para llevar a cabo este proceso

|   |   |
|---|---|
| <br>UNIVERSIDAD PEDAGÓGICA<br>NACIONAL<br><small>Formación de Maestros</small> | <b>FORMATO</b>                              |
|   | <b>RESUMEN ANALÍTICO EN EDUCACIÓN - RAE</b> |
| <b>Código: FOR020GIB</b>  | <b>Versión: 01</b>                          |
| <b>Fecha de Aprobación: 10-10-2012</b>  | <b>Página 4 de 4</b>                        |

fue necesario realizar un ajuste y validación a algunas categorías, ya que varias de las respuestas de los estudiantes no se lograban ubicar con claridad en una categoría específica. Se considera que el ajuste realizado a dichas categorías, es un aporte de este trabajo que mejora y hace más clara la interpretación de respuestas de los estudiantes, en el sentido de facilitar la identificación de las fortalezas, dificultades y falencias que se presentan en el proceso de enseñanza – aprendizaje de los gráficos de caja.

El cuestionario de evaluación propuesto por Palacios y Zambrano (2009) requiere de algunos ajustes que le permitan al estudiante observar con claridad la información presentada, como es el caso de los gráficos de caja que representan las dos situaciones con las que se caracteriza el razonamiento estadístico de los estudiantes.

En el presente estudio se observa que los estudiantes de grado noveno de la IED Técnico Comercial Mariano Ospina Rodríguez, presentan un mejor nivel de razonamiento estadístico en la mayoría de los elementos de análisis establecidos, comparado con los estudiantes de grado undécimo de la IED General Carlos Albán, que fueron los dos grupos de estudio considerados. Esta situación parece coherente con el hecho de observar un mayor nivel académico de la IED Técnico Comercial Mariano Ospina Rodríguez, registrado según el ICFES en un nivel alto, en comparación con el registrado por la IED General Carlos Albán, registrado en un nivel medio. Por lo tanto no se confirmó el supuesto implícito, de que a mayor nivel escolar, mayor nivel de razonamiento estadístico.

De acuerdo al análisis que se llevó a cabo, los elementos de razonamiento estadístico que reflejan un nivel idiosincrásico de los estudiantes analizados en el presente proyecto en su orden son: Tamaño de muestra, localización, intervalos de concentración, escala, dispersión, datos atípicos y datos típicos extremos. Mientras que para la Población de estudio del proyecto Palacios y Zambrano (2009) son en su orden: Dispersión, Localización, Intervalos de Concentración, Tamaño de la muestra, Datos adyacentes (datos típicos extremos), Escala y Datos Atípicos.

El bajo nivel de razonamiento estadístico de los estudiantes puede ser atribuido a que no han recibido educación estadística constante, siendo para algunos de ellos la aplicación de las guías propuestas un primer acercamiento al conocimiento de la estadística.

|                       |  |
|-----------------------|--|
| <b>Elaborado por:</b> | CHAPARRO, Edilma; ROMERO, José Alcides |
| <b>Revisado por:</b>  | FERNÁNDEZ HERNÁNDEZ, Jorge Felipe      |

|  |    |    |      |
|--|----|----|------|
| <b>Fecha de elaboración del Resumen:</b> | 28 | 10 | 2012 |
|--|----|----|------|

# TABLA DE CONTENIDO

|   |    |
|---|----|
| <b>TABLA DE CONTENIDO</b> .....   | 10 |
| <b>LISTA DE GRÁFICOS</b> .....  | 12 |
| <b>LISTA DE TABLAS</b> .....  | 13 |
| <b>INTRODUCCIÓN</b> .....   | 14 |
| <b>1. PRELIMINARES</b> .....  | 16 |
| JUSTIFICACIÓN.....  | 16 |
| ANTECEDENTES.....   | 17 |
| PLANTEAMIENTO DEL PROBLEMA.....   | 17 |
| OBJETIVOS.....  | 18 |
| Objetivo General.....   | 18 |
| Objetivos Específicos.....  | 18 |
| <b>2. MARCO CONCEPTUAL</b> .....  | 19 |
| ANÁLISIS DE DATOS.....  | 19 |
| Análisis Exploratorio de Datos.....   | 20 |
| El gráfico de caja.....   | 21 |
| Gráfico tipo 1.....   | 22 |
| Gráfico tipo 2.....   | 23 |
| Gráfico tipo 3.....   | 24 |
| Elementos de análisis de un gráfico de cajas.....   | 24 |
| RAZONAMIENTO ESTADÍSTICO.....   | 25 |
| Razonamiento estadístico en la educación estadística.....                                 | 25 |
| Razonamiento al comparar conjuntos de datos.....  | 26 |
| Taxonomía de la Estructura de los Resultados Observados en el Aprendizaje (SOLO)<br>..... | 27 |
| PROPUESTA DE INSTRUCCIÓN.....   | 30 |
| <b>3. ENFOQUE METODOLÓGICO</b> .....  | 33 |
| FASE I. ANÁLISIS PRELIMINARES Y PROPUESTA DE INSTRUCCIÓN.....                             | 33 |
| Decisiones acerca del gráfico de caja.....  | 33 |
| Categorías de análisis planteadas por Palacios y Zambrano.....                            | 34 |
| Propuesta de instrucción.....   | 34 |
| FASE II. IMPLEMENTACIÓN Y RECOLECCIÓN DE INFORMACIÓN.....                                 | 35 |
| Muestra de estudio.....   | 35 |

|  |    |
|--|----|
| Implementación del instrumento de evaluación .....                     | 35 |
| FASE III ANÁLISIS DE RESULTADOS .....                                  | 36 |
| Revisión y ajuste de las categorías.....                               | 36 |
| Categorías ajustadas (Validación de categorías) .....                  | 36 |
| Sistematización y análisis de resultados.....                          | 36 |
| Comparación de poblaciones.....  | 36 |
| 4. RESULTADOS .....  | 38 |
| CLASIFICACIÓN DE LAS RESPUESTAS POR ÍTEMS .....                        | 38 |
| TIPOS DE RESPUESTAS DE LOS ESTUDIANTES POR CATEGORÍAS .....            | 52 |
| PRESENTACIÓN Y ANÁLISIS DE RESULTADOS.....                             | 61 |
| <b>Con respecto al Tamaño de la Muestra</b> .....                      | 61 |
| <b>Con respecto a la Escala</b> .....                                  | 62 |
| <b>Con respecto a la Localización</b> .....                            | 62 |
| <b>Con respecto a la Dispersión</b> .....                              | 63 |
| <b>Con respecto a los Intervalos</b> .....                             | 64 |
| <b>Con respecto a los datos típicos extremos</b> .....                 | 64 |
| <b>Con respecto a los Datos Atípicos</b> .....                         | 65 |
| <b>Con respecto al Total</b> .....                                     | 66 |
| 5. CONCLUSIONES.....   | 67 |
| <b>RESPECTO A LOS OBJETIVOS DEL ESTUDIO</b> .....                      | 67 |
| <b>RESPECTO A LA PROPUESTA DE INSTRUCCIÓN</b> .....                    | 67 |
| <b>RESPECTO A LOS NIVELES DE CATEGORIZACIÓN DEL RAZONAMIENTO</b> ..... | 68 |
| <b>RESPECTO AL INSTRUMENTO DE EVALUACIÓN</b> .....                     | 68 |
| <b>RESPECTO A LOS RESULTADOS</b> .....                                 | 69 |
| REFERENCIAS .....  | 71 |
| ANEXOS.....  | 73 |

## LISTA DE GRÁFICOS

|   |    |
|---|----|
| Gráfico 1. Gráfico de cajas con Límites ubicados a tres recorridos intercuartiles ..... | 23 |
| Gráfico 2. Gráfico de cajas con Límites ubicados a 1.5 recorridos intercuartiles .....  | 23 |
| Gráfico 3. Gráfico de cajas sin zonas de anomalías .....                                | 24 |
| Gráfico 4. Situación uno del instrumento de evaluación .....                            | 38 |
| Gráfico 5. Situación dos del instrumento de evaluación .....                            | 45 |
| Gráfico 6. Tamaño de muestra .....  | 61 |
| Gráfico 7. Escala .....   | 62 |
| Gráfico 8. Localización .....   | 62 |
| Gráfico 9. Dispersión .....   | 63 |
| Gráfico 10. Intervalos de Concentración .....   | 64 |
| Gráfico 11. Datos Típicos Extremos .....  | 64 |
| Gráfico 12. Datos Atípicos .....  | 65 |
| Gráfico 13. Análisis global .....   | 66 |

## LISTA DE TABLAS

|  |    |
|--|----|
| Tabla 1. Categorías de razonamiento estadístico vs ítems del cuestionario Palacios y Zambrano (2009) ..... | 38 |
| Tabla 2. Categorías de razonamiento estadístico Palacios y Zambrano (2009) .....                           | 54 |
| Tabla 3. Categorías de razonamiento estadístico modificada .....   | 56 |
| Tabla 4. Respuestas de los estudiantes por categoría y nivel .....   | 60 |

# INTRODUCCIÓN

La educación estadística cada día ha tenido una mayor acogida en la comunidad educativa, debido a la cantidad de información que se observa a través de diferentes medios de comunicación, la necesidad constante de aprender a ser críticos sobre la realidad que se muestra y además formar estudiantes capaces de enfrentarse a las diferentes pruebas propuestas por el Ministerio de Educación Nacional, como Internacional. El MEN propone el análisis de datos como una manera de acercar más al estudiante al razonamiento de un conjunto de datos que le permita interpretar y analizar sin tratarse únicamente de la aplicación de fórmulas; es así como el Análisis Exploratorio de Datos (AED) que fue introducido por Tukey (1977) adquiere vital importancia por sus diferentes formas gráficas de representar un conjunto de datos; entre ellas se encuentra el gráfico de caja, que permite observar medidas de centralidad, localización, dispersión, datos inusuales, como también llevar a cabo comparaciones de dos o más conjuntos de datos de forma visual.

De esta manera la presente propuesta tiene como objetivo “Promover en los estudiantes de grado noveno y undécimo comprensión y razonamiento estadístico al comparar distribuciones de datos representados en gráficos de caja” planteando una serie de actividades que aborden el aprendizaje de este gráfico, para posteriormente ser caracterizado en diferentes niveles de razonamiento, teniendo en cuenta una población de 91 estudiantes de los grados noveno y undécimo de las IED Técnico Comercial Mariano Ospina Rodríguez de Guasca Cundinamarca y General Carlos Albán de Albán Cundinamarca. Además, del interés de verificar un proyecto de maestría realizado por Palacios y Zambrano (2009), en el cual se caracteriza el razonamiento estadístico de los estudiantes, por ésta razón el presente trabajo aborda los elementos y niveles de razonamiento que proponen los autores de la tesis. Se llevo a cabo mediante tres fases, en la primera se realizó la lectura y revisión de la tesis mencionada, literatura relacionada con el Análisis Exploratorio de Datos, niveles de razonamiento basados en la taxonomía SOLO y propuestas de enseñanza de los gráficos de caja. En la segunda fase se diseñó e implementó una secuencia de actividades para la enseñanza del gráfico de caja, valorando el desempeño de los estudiantes mediante el instrumento de evaluación propuesto por Palacios y Zambrano (2009). En la tercera fase se realizó la categorización del razonamiento estadístico de los estudiantes, reajustando las categorías de análisis,

sistematizando la información recolectada y comparando los resultados obtenidos por la población de estudio del trabajo de maestría, con el total del grupo de estudiantes del presente trabajo a la vez discriminando la población de los grados noveno y undécimo. Finalmente se plantean conclusiones acerca del alcance de los objetivos propuestos, las actividades diseñadas para la enseñanza del gráfico de caja, los niveles de caracterización del razonamiento estadístico, del instrumento de evaluación y los resultados obtenidos.



# 1. PRELIMINARES

## JUSTIFICACIÓN

La necesidad que tienen las personas del común o especialistas en diferentes campos, de comprender y analizar información de datos que se presenta en distintos medios y formas de representación, es uno de los asuntos que debe considerar la Educación Estadística como disciplina y área de trabajo . De hecho, hoy es muy común que en los currículos escolares de muchos países se le vea como parte esencial de la formación de un ciudadano.

Por ejemplo, en el caso de Colombia, el Ministerio de Educación Nacional [MEN] (2006) propone mediante los Estándares Básicos de Competencias en Matemáticas que en el plan de estudios de las Instituciones Educativas se incluya el llamado “Pensamiento aleatorio y sistemas de datos”. A través de la conceptualización de este tipo de pensamiento se busca sugerir y plantear recomendaciones para que el profesor promueva en el estudiante la apropiación de destrezas en la recolección, organización e interpretación de resultados así como para la formulación de conjeturas, basada en evidencia estadística.

Día a día se hace más evidente la necesidad que las personas adquieran conocimientos estadísticos que les permitan reconocer, interpretar y criticar la información presentada por diferentes medios; es así como en el ámbito escolar es de vital importancia fortalecer la cultura estadística, teniendo en cuenta que es una competencia valorada en las pruebas ICFES, los docentes de las instituciones educativas deben procurar plantear situaciones que fomenten en los estudiantes el razonamiento estadístico.

En las Instituciones Educativas “Mariano Ospina Rodríguez de Guasca” y “General Carlos Albán de Albán” la enseñanza de la estadística se ha dejado en cierto modo de lado, puesto que esta se propone como último tema dentro de la planeación del área de matemáticas o como asignatura en el grado undécimo con una intensidad semanal de una hora. De esta manera, el presente trabajo ha considerado el estudio del razonamiento estadístico de los estudiantes de grado noveno y undécimo de las Instituciones educativas

mencionadas, cuando se aborda la interpretación y comparación de información representada mediante gráficos de cajas.

## **ANTECEDENTES**

El presente proyecto toma como principal antecedente el trabajo desarrollado por Palacios y Zambrano (2009). La idea es continuar el proceso de la investigación desarrollado por ellos en el programa de Maestría de Docencia de las Matemáticas de la Universidad Pedagógica Nacional. El trabajo de grado de estos estudiantes fue titulado “El Razonamiento estadístico de estudiantes de secundaria al comparar conjuntos de datos representados mediante gráficos de caja” Palacios y Zambrano (2009).

Otros trabajos que sirven de antecedentes, citados también en el trabajo de Palacios y Zambrano (2009) que fue necesario consultar para complementar la conceptualización de Razonamiento Estadístico fueron revisados Pfannkuch (2006), Wild y Pfannkuch (1999), Garfield (2002) y Ben - Zvi (2004). Por otra parte, para complementar lo que tiene que ver con Modelos de Razonamiento Estadístico se revisó los trabajo de Jones et al. (2004).

Y en relación con las características del Análisis Exploratorio de Datos y en particular lo relativo al grafico de caja se consultó Velleman y Hoaglin (1981).

## **PLANTEAMIENTO DEL PROBLEMA**

La información estadística se puede representar a través de diferentes gráficas, algunas comúnmente usadas y enseñadas como: el diagrama de barras, diagrama circular, histogramas, series de tiempo, entre otras. Una representación que permite visualizar fácilmente la información y potencializa el razonamiento estadístico de los estudiantes es el diagrama de caja, que se ha venido introduciendo en los últimos años en el análisis exploratorio de datos y cuya enseñanza no es muy común en la secundaria.

Este gráfico también denominado *Box plot*, fue introducido por Tukey (1977). Es una representación gráfica que proporciona los valores de los datos pero sin que estos aparezcan con todo detalle, a fin de no perder su configuración espacial; para ello mantiene únicamente sus valores indicativos principales, que pueden resumirse en: Localizaciones y agrupaciones significativas de valores, zonas en las que predomina la dispersión, referencia visual de la simetría central y de los extremos, rango, valores atípicos, anomalías o valores significativamente alejados.

Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos, a saber: mediana, primer cuartil, tercer cuartil y dos valores extremos: el valor máximo y el valor mínimo; además presenta al mismo tiempo información sobre la tendencia central, la dispersión y la simetría de un conjunto de datos permitiendo identificar con claridad y de forma individual observaciones que se alejan de manera poco usual del resto de los datos; a estas observaciones se les llama valores atípicos. Admite igualmente comparar a la vez varios grupos de datos sin perder información.

De esta manera, el interés principal del presente trabajo es el de fomentar la comprensión y el razonamiento estadístico de los estudiantes del grado noveno y undécimo al interpretar información representada en gráfico de cajas. Se espera que, en consonancia con lo sugerido por el MEN (2006), este trabajo aporte elementos para el diseño y aplicación de propuestas curriculares que contribuyan en la enseñanza y aprendizaje de la estadística. Por ello, se considera importante dar respuesta a la siguiente pregunta de investigación: ¿Cómo promover el desarrollo del razonamiento estadístico de los estudiantes de grado noveno y undécimo a través del análisis de distribuciones de datos representados por gráficos de caja?

## **OBJETIVOS**

### **Objetivo General**

Promover en los estudiantes de grado noveno y undécimo comprensión y razonamiento estadístico al comparar distribuciones de datos representados en gráficos de caja.

### **Objetivos Específicos**

- Plantear una serie de actividades que aborde el aprendizaje y la comprensión de información representada en gráficos de caja.
- Identificar y describir elementos que caracterizan el razonamiento estadístico subyacente a la comprensión de un gráfico de caja.
- Analizar el razonamiento estadístico de los estudiantes de grado noveno y undécimo al comparar distribuciones de datos representados por gráficos de caja.

## **2. MARCO CONCEPTUAL**

Este capítulo se precisa la base conceptual que fundamenta el desarrollo del trabajo a partir de los antecedentes de los trabajos mencionadas en el capítulo anterior y en particular basados en la propuesta de Zambrano y Palacios (2009). El desarrollo de esta marco se divide en tres partes: La primera hace relación al análisis de datos, destacando el análisis exploratorio de datos (AED) e identificando el objeto de estudio y los elementos que lo conforman, como son: origen, construcción y estructura, además de los diferentes tipos de gráficos de caja.

La segunda relaciona el razonamiento estadístico al comparar dos conjuntos de datos, de igual manera presenta los diferentes elementos y niveles de razonamiento estadístico que se tienen en cuenta para la valoración de los estudiantes, basados en la Taxonomía SOLO.

Finalmente, en la tercera parte se mencionan los lineamientos que guían la estructura de elaboración de la propuesta de enseñanza de los gráficos de caja.

### **ANÁLISIS DE DATOS**

La búsqueda de respuestas a preguntas sobre el mundo físico resulta ser una actividad llena de sentido para los estudiantes, si se hace a través de recolección y análisis de datos que permita decidir la pertinencia de la información necesaria, la forma de recogerla, de representarla y poner en práctica los conocimientos para realizar mediciones, estimaciones y plantear estrategias de resolución de problemas matemáticos y de otras áreas del conocimiento.

El análisis de datos es una rama importante de la Estadística que ha evolucionado desde manejos elementales en el tratamiento de la información, hasta el uso cada día más común de software especializados. Al respecto Palacios y Zambrano, citando a Cobo (2003) plantean: “Puesto que los ordenadores permiten analizar grandes muestras y gran número de variables con rapidez y fiabilidad, ya no hay que limitarse a los métodos estadísticos basados en distribuciones conocidas, cuya principal aplicación eran las pequeñas muestras”, si no que se puede realizar un análisis descriptivo amplio y en menor tiempo. Además de recolectar los datos, organizarlos y sistematizarlos permitiría

plantear métodos que faciliten la toma de decisiones o estimar características de una población a partir de la información que se tiene de una muestra, es decir realizar inferencias de la población a partir de ésta. Según Palacios y Zambrano (2009) “el planteamiento de inferencias se basa en la teoría de probabilidad, pues a diferencia del análisis descriptivo, el inferencial permite considerar aspectos aleatorios muy comunes en situaciones cotidianas, donde casi siempre está presente el azar y no es posible obtener los datos de una población completa”.

### **Análisis Exploratorio de Datos**

El Análisis Exploratorio de Datos (AED) emplea una variedad de técnicas gráficas para explorar distribuciones de datos. Este grupo de técnicas gráficas y analíticas creado por el estadístico norteamericano Tukey (1977), permite conseguir un conocimiento previo de los datos a analizar siempre desde una perspectiva exploratoria, considerándose como una nueva rama de la estadística; Sin embargo, algunos autores como Bakker (2004) sostienen que “a diferencia de la estadística descriptiva tradicional el objetivo en el AED no es probar una hipótesis preestablecida con datos de un muestreo aleatorio, sino más bien éste se enfoca en la dirección de patrones y tendencias no anticipadas de todo tipo de datos, ya sea una muestra aleatoria o no”.

Las técnicas estadísticas empleadas por el AED, buscan favorecer el estudio de estadísticos robustos que permitan observar la estructura de un conjunto de datos sin que esta se vea demasiado afectada por la presencia de datos atípicos. En ocasiones se selecciona una muestra donde se nota la presencia de alguno de estos datos inusitados y es necesario establecer si tal observación hace parte importante de la muestra y si se debe retener o rechazar, pues la presencia de este dato hace que las medidas descriptivas que generalmente se calculan, pueden verse afectadas y por ejemplo, la media aritmética ya no sea una medida representativa del conjunto de datos.

Por otro lado, la importancia del AED es resaltada por Ben-Zvi (2000), quien sugiere que: “El análisis exploratorio de datos es la disciplina de organización, descripción, representación y análisis de datos, con una fuerte confianza en las herramientas analíticas y visuales. Su objetivo principal es dar sentido y buscar más allá de los datos para que, de esta manera, junto a la inferencia, se puedan explorar nuevos datos.” Así

esta nueva filosofía no se limita a la construcción de gráficos estadísticos, aunque los dos términos se usan casi alternativamente.

Los gráficos estadísticos son una colección de técnicas que se centran en un aspecto de la caracterización de los datos que son a menudo absolutamente simples, consistentes en el diseño de diagramas para descubrir algunos aspectos que no son apreciables a simple vista. En términos generales la meta fundamental del AED es mejorar la inserción del observador en un conjunto de datos y en la estructura subyacente, mientras que proporciona todos los elementos necesarios que se desearía extraer de un conjunto de datos. Aunque conseguir una descripción de los datos no es suficiente para que el analista sepa qué sucede con ellos, a diferencia de la estadística descriptiva, el AED le concede la misma importancia a los dos componentes que conforman los datos: la centralidad y la dispersión.

Por otro lado “Las posibilidades didácticas del análisis exploratorio de datos se deben a la sencillez del aparato matemático requerido, la importancia dada hoy día en estadística y matemáticas a los sistemas de representación múltiple y resolución de problemas, las conexiones con otros temas del currículo, el trabajo en equipo y la posibilidad de desarrollo de proyectos por parte de los alumnos” Batanero, Estepa y Godino (1991).

Respecto a las representaciones gráficas y fortalezas del AED y al enfoque casi exclusivamente algorítmico de la estadística descriptiva, Freixa (1992), sostiene que: “sin duda alguna, la principal novedad del AED es la forma de presentar un conjunto de datos en representaciones visuales o gráficas como, por ejemplo, el diagrama de tallo y hojas o el gráfico de cajas. Este tipo de representación no solo permite descubrir los patrones específicos o de tendencia, sino que a su vez ponen de manifiesto aspectos sorprendentes, insospechados o a veces divertidos, que de otra forma pasarían totalmente inadvertidos. Junto a estos procedimientos de representación visual deben también destacarse, por su utilidad en la descripción más precisa de las formas o patrones que toman los datos, los gráficos de valores de letra o los gráficos de Caja, entre otros”.

### **El gráfico de caja**

El gráfico de caja es una herramienta del análisis exploratorio de datos, que permite visualizar fácilmente la información proveniente de uno o varios conjuntos de datos. El

nombre original es Box and whisker plot, es decir, gráfico de caja y bigotes fue introducido por Tukey (1977) en el segundo capítulo de su libro Exploratory Data Analysis. Es un gráfico representativo de las distribuciones de un conjunto de datos en cuya construcción se usan cinco medidas descriptivas de los mismos, a saber: mediana, primer cuartil, tercer cuartil y dos valores extremos: el valor máximo y el valor mínimo; además permite visualizar información sobre la tendencia central, la dispersión y la simetría de un conjunto de datos, favoreciendo identificar con claridad y de forma individual observaciones que se alejan y sobrepasan los valores adyacentes; a estas observaciones se les llama datos atípicos que influyen significativamente en el cálculo de las medidas que caracterizan la variable, a la vez admite comparar varios grupos de datos sin perder información.

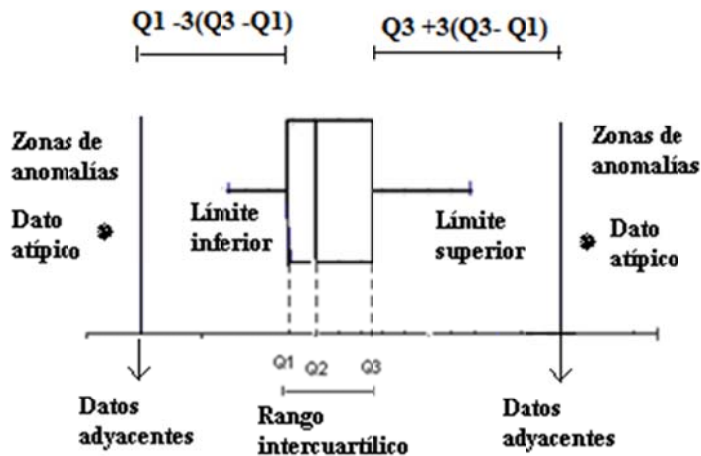
Esta representación gráfica proporciona los valores de los datos, sin que estos aparezcan con todo detalle, a fin de no perder su configuración espacial; para ello mantiene únicamente valores indicativos principales, que pueden resumirse en:

- Localizaciones y agrupaciones significativas de valores
- Zonas en las que predomina la dispersión
- Referencia visual de la simetría central y de los extremos
- Rango
- Datos adyacentes, atípicos, anomalías o valores significativamente alejados.

A través de la revisión documental acerca del objeto de estudio, se encontraron tres tipos de gráficos de caja:

#### *Gráfico tipo 1*

El gráfico que se clasificó como tipo uno fue propuesto por Velleman (1981), quién plantea los datos adyacentes como límites de las zonas de anomalías, teniendo en cuenta que para identificar el dato adyacente menor se hace uso de los cuartiles  $Q3 + 3(Q3 - Q1)$  y para el mayor  $Q1 - 3(Q3 - Q1)$ , y a partir de estos extremos se encuentran los datos atípicos. Ver grafico 1.



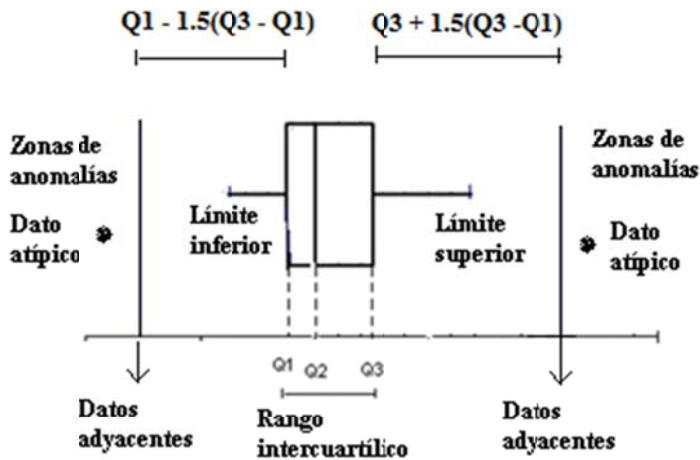
**GRÁFICO 1**

Gráfico 1. Gráfico de cajas con Límites ubicados a tres recorridos intercuartiles

*Gráfico tipo 2*

El gráfico tipo dos es considerado por Zambrano y Palacios (2009) en su tesis de maestría, se diferencia del tipo uno, en las cercas que limitan las zonas de anomalías, ya que se trazan a partir de la suma o diferencia del cuartil 1 (Q1) y cuartil 3 (Q3), teniendo en cuenta uno punto cinco veces el rango intercuartílico.

( $Q3 + 1.5(Q3 - Q1)$  y  $Q1 - 1.5(Q3 - Q1)$ ). Ver gráfico 2



**GRÁFICO 2**

Gráfico 2. Gráfico de cajas con Límites ubicados a 1.5 recorridos intercuartiles



### Gráfico tipo 3

El siguiente gráfico considerado como tipo tres es el más comúnmente encontrado en los libros de texto y algunos artículos de estadística, se puede observar que es un gráfico que no tiene en cuenta la presencia de datos atípicos en la distribución y los bigotes son prolongados hasta los datos mínimo y máximo.

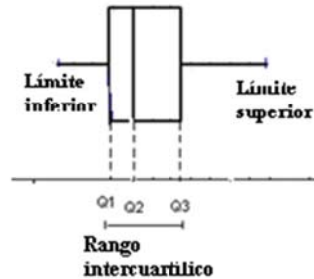


GRÁFICO 3

Gráfico 3. Gráfico de cajas sin zonas de anomalías

### Elementos de análisis de un gráfico de cajas

Para caracterizar el razonamiento estadístico de los estudiantes se tuvieron en cuenta los siguientes elementos de análisis que se pueden identificar en un gráfico de cajas:

#### Tamaño de la muestra:

En el gráfico de cajas se pueden observar ciertos intervalos de concentración de los datos, valores adyacentes y atípicos, límites superior e inferior al igual que el valor de los cuartiles; sin embargo no es posible identificar el tamaño de la muestra. Frente a este elemento se espera que el estudiante no relacione el número de datos (Tamaño de la muestra) con la forma o el tamaño del gráfico. De tal manera que empleando únicamente el gráfico de cajas no se puede determinar la cantidad de datos analizados.

**Escala:** El gráfico de cajas emplea la recta numérica como referencia para ubicar los cinco estadísticos resumen de tal manera que su adecuado empleo permite visualizar de manera correcta su ubicación y mostrar la dispersión de los datos. Además en su construcción se debe hacer un uso correcto de las escalas para evitar errores de interpretación en su lectura.

**Localización:** En un gráfico de cajas se considera como la capacidad que tiene el estudiante de ubicar e interpretar los valores que representan los cinco resúmenes (Valor Mínimo, el Cuartil 1, La Mediana, el Cuartil 3 y el valor máximo) y los datos adyacentes; además de ubicar los límites de los intervalos de concentración de los datos y a su vez relacionarlos con el fin de establecer conjeturas acerca de la distribución de los datos.

**Dispersión:** A través de los gráficos de cajas se puede establecer si la población a analizar es o no homogénea, mediante el cálculo de la distancia entre los diferentes intervalos de concentración del conjunto de datos que se visualiza de manera gráfica con la amplitud de la caja y el tamaño de los bigotes, guardando directa proporción a la dispersión de los datos.

**Intervalos de concentración:** Se refiere a los agrupamientos de los datos que se establecen entre los valores de los cinco resúmenes, indicando entre estos la tendencia de los datos y los sesgos de la población. Este elemento de análisis permite describir la forma de la distribución de los datos; a través de estos intervalos se puede establecer si estos presentan sesgos (Izquierda o Derecha), además de poder establecer el grado de concentración de los datos con respecto a los valores centrales (Distribución puntiaguda, en forma de campana o plana)

**Datos típicos extremos:** Considerados como los valores correspondiente a los límites de los dos bigotes dentro de la zona típica determinada por los datos adyacentes planteados por Velleman (1981).

**Datos atípicos:** Son aquellos datos que sobrepasan los valores adyacentes y se ubican en la zona de anomalías. Para su análisis se espera que los estudiantes los identifiquen como aquellos valores (ya sean superiores o inferiores) que se salen del común de los datos. El razonamiento acerca de estos valores se hace en torno a cómo afecta su presencia o no en la toma de decisiones.

## **RAZONAMIENTO ESTADÍSTICO**

### **Razonamiento estadístico en la educación estadística**

Siendo la educación estadística de vital importancia para los ciudadanos, la comunidad investigativa de esta área, ha estado muy interesada en encontrar diferencias en

definiciones que se hacen acerca de esta, como son: instrucción estadística, razonamiento y pensamiento estadístico pero aún no se ha hecho ningún acuerdo formal con respecto a las definiciones y distinciones de estas. Sin embargo Ben-Zvi y Garfield (2004) proponen que: La *Instrucción estadística* incluye las habilidades básicas e importantes que se pueden utilizar con base en la información de los resultados estadísticos; estas habilidades incluyen organizar las tablas datos y presentar los resultados mediante diversas representaciones, igualmente comprender vocabulario, simbología y comprender la probabilidad como medida de incertidumbre. Por otro lado el *Razonamiento estadístico* puede ser definido como las interpretaciones que realizan las personas basadas en sistemas, representaciones o resúmenes estadísticos de datos. El razonamiento estadístico implica conectar un concepto a otro o combinar ideas sobre datos que permitan explicar procesos estadísticos e interpretar completamente los resultados estadísticos. De igual modo, el *pensamiento estadístico* implica una comprensión del porqué y del cómo es que se conducen las investigaciones. Estas ideas incluyen la naturaleza de cómo y cuáles son los métodos apropiados para el análisis de datos. El pensamiento estadístico implica comprensión de la naturaleza del muestreo, de cómo se hace inferencias a partir de una muestra para generalizar en la población. Incluye una comprensión de los modelos a utilizar para simular datos, estimar probabilidades y valorar herramientas deductivas existentes pueden ser utilizadas por el investigador de manera crítica para ayudar al proceso y evaluar los resultados de un problema solucionado o de un estudio estadístico.

### **Razonamiento al comparar conjuntos de datos**

En este apartado se describe la investigación de Makar y Confrey citada por Palacios y Zambrano (2009), quienes estudiaron el razonamiento estadístico de docentes al comparar conjuntos de datos. Este estudio fue llevado a cabo al final de una secuencia de desarrollo profesional de 6 meses como parte de un proyecto de investigación NSF- en la Universidad de Texas de Austin. El trabajo de estos autores se enfoca en el razonamiento estadístico de cuatro profesores de secundaria a lo largo de entrevistas dirigidas al final de una secuencia de desarrollo profesional. Los profesores dirigieron investigaciones usando el software Fathom, al plantear el siguiente problema de investigación: "¿cómo se determina si dos grupos son distintos? Mediante un análisis cualitativo se examinaron las respuestas como resultado de las entrevistas en las que se les pidió a los profesores que describieran el desempeño relativo de dos grupos de estudiantes en las pruebas de

estado de matemáticas. El análisis cuantitativo previo y el posterior sobre el conocimiento estadístico proporcionaron la triangulación necesaria para un conocimiento posterior sobre la comprensión de los profesores.

Aunque este estudio se desarrolló en torno a la descripción del razonamiento estadístico de profesores de secundaria, el objetivo de los autores fue ofrecer la posibilidad de entender el razonamiento estadístico acerca de la variación en contextos más sofisticados

Los autores con esta investigación concluyeron entre otras cosas que en el nivel de enseñanza elemental, la comparación de dos conjuntos de datos puede ser usada para introducir conceptos del análisis de datos al proveer experiencias que faciliten la visualización y el razonamiento acerca de distribuciones y que en la educación media ya es posible integrar estas experiencias con datos al manejo de medidas de centralidad, de dispersión y al reconocimiento de la forma de cada distribución y de datos atípicos. Además, afirman que los estudiantes de grados superiores pueden ser introducidos en la inferencia estadística informal y en la toma de decisiones a partir de los datos.

### **Taxonomía de la Estructura de los Resultados Observados en el Aprendizaje (SOLO)**

La taxonomía SOLO es un sistema de categorías diseñado para evaluar la calidad de una respuesta Peeg, Gutiérrez y Huerta, (1997). Su origen se puede encontrar en formulaciones piagetianas y reformulaciones neo piagetianas posteriores Biggs y Collis (1991) de las etapas del desarrollo cognitivo de Piaget. Dos aspectos deben tenerse en cuenta cuando se usa la taxonomía SOLO: Los modos de funcionamiento (niveles de abstracción que van progresando desde las acciones concretas a los principios y conceptos abstractos, lo que forma la base de las etapas evolutivas) y los ciclos del aprendizaje Biggs y Collis (1991). El primero está fuertemente relacionado con las etapas piagetianas del desarrollo cognitivo. El segundo, tiene que ver con describir la estructura de cualquier respuesta como un fenómeno en sí mismo, esto es, sin que la respuesta represente necesariamente una etapa particular en el desarrollo intelectual. Este segundo aspecto es la esencia de la taxonomía. Este es modelo que puede ser una valiosa herramienta para analizar la estructura de las respuestas escritas presentadas por los estudiantes, identificando el tipo de razonamiento demostrado en éstas, para lo cual

establece algunas jerarquías en las concepciones de los estudiantes sobre un determinado objeto de conocimiento.

Los niveles de la Taxonomía SOLO se pueden clasificar en dos categorías mayores, „Superficial y Profunda, las cuales a su vez contienen dos niveles. En la primera categoría, es decir, la superficial se pueden distinguir los niveles Uniestructural y Multiestructural, mientras que la categoría Profunda está compuesta por el nivel Relacional y el Abstracto Extendido. Sin embargo, también se considera un nivel para aquellas respuestas que evidencian un total desconocimiento del concepto evaluado llamado nivel Idiosincrásico o Preestructural.

Estos niveles están ordenados en términos de varias características: en primer lugar, desde lo concreto a lo abstracto; en segundo lugar, al considerar un creciente número de dimensiones de organización; y por último, teniendo en cuenta la consistencia en los resultados cualitativos en el aprendizaje. A continuación se describen estos cinco niveles de desarrollo, especificando además algunas acciones relevantes que indican lo que un estudiante podría hacer en cada nivel.

**Nivel Idiosincrásico o Preestructural:** Las respuestas de los estudiantes en este nivel se caracterizan por no demostrar ninguna aproximación a la posibilidad de hacer conexiones entre los componentes de la información que se presenta. Generalmente pueden contestar con un simple “no lo sé” o con lo primero que se le viene a la mente fallando en la pregunta.

**Nivel Uniestructural:** Las respuestas de los estudiantes se enfocan sobre sólo un aspecto del concepto evaluado, sin hacer ningún tipo de conexión con otras ideas, pero es capaz de razonar sobre algún contenido significativo. Se caracteriza por acciones como: identificar, memorizar o realizar procedimientos simples.

**Nivel Multiestructural:** En este nivel las respuestas de los estudiantes identifican más de un aspecto del concepto evaluado, pero no determinan el significado del todo pues las ideas involucradas están desordenadas y sin relación. Dos respuestas a este nivel pueden tomar la misma información, pero no es posible establecer conexiones por lo que llevan a conclusiones diferentes. Se caracteriza por acciones como: enumerar, clasificar, describir, combinar y efectuar algoritmos.

**Nivel Relacional:** Los estudiantes a este nivel pueden indicar conexiones entre los hechos y la teoría, las acciones y los propósitos. Las respuestas muestran cierta comprensión de algunos componentes, los cuales son integrados conceptualmente, demostrando cómo las partes contribuyen al todo para darle estructura y significado. Sin embargo, las respuestas dependen de experiencias concretas o de los conceptos enseñados anteriormente, por lo que el estudiante puede aplicar el concepto a problemas familiares. Se caracteriza por acciones como: comparar, contrastar, explicar causas, integrar, analizar y relacionar.

**Nivel Abstracto Extendido:** Las respuestas a este nivel consideran toda la información relevante y relacionan sus componentes bajo una estructura hipotética abstracta que hace posible realizar deducciones acerca de información que no fue incluida originalmente. Además, vincula conceptos que no han sido explícitamente enseñados. La comprensión es transferible y generalizable a otras situaciones o a otras áreas. Se caracteriza por acciones como: teorizar, generalizar, conjeturar y reflexionar.

Sin embargo, en algunas ocasiones las respuestas de los estudiantes no se ajustan exactamente en ninguno de los cinco niveles anteriormente mencionados. De allí que Biggs y Collis (1982) se refirieron a éstas como respuestas de transición que se dan cuando la respuesta del estudiante está próxima a un determinado nivel sin ubicarse en este. A menudo se manifiesta por una determinada confusión cuando el estudiante está manipulando más información de la que puede y no presenta argumentos sólidos.

Además, en concordancia con el modelo general de desarrollo de Biggs y Collis (1982), el razonamiento estadístico de los estudiantes, se puede caracterizar como un proceso en desarrollo a través de niveles que reflejan cambios en su complejidad. De esta manera, Jones y Mooney citado por Palacios y Zambrano (2009) describen el razonamiento estadístico con respecto a los cuatro procesos y los caracterizan a través de los niveles ya mencionados. Este marco provee descriptores específicos acerca del *razonamiento estadístico* de los estudiantes en relación a los procesos para cada nivel, los cuales son:

- *Describir datos:* Este proceso involucra la lectura explícita de conjuntos de datos presentados por tablas, gráficos. Generando la posibilidad de caracterizar dos subprocesos importantes: la conciencia de representar características importantes e identificar unidades de valores de datos.

- *Organizar datos:* Este proceso involucra la organización de los datos en grupos o conglomerados que a su vez los lleva a descubrir patrones o tendencias, categorizar y consolidar datos a manera de resumen. En este sentido, es importante el uso de medidas descriptivas como las de tendencia central o de dispersión para comparar conjuntos de datos. Es posible caracterizar tres subprocesos importantes: agrupar datos, resumir datos en términos de la centralidad y describir la dispersión de un conjunto de datos.
- *Representar datos:* Este proceso involucra describir y analizar datos de forma gráfica. Conllevando a caracterizar dos subprocesos importantes: completar o construir la representación de un conjunto de datos y evaluar la eficacia de alguna representación de datos.
- *Analizar e interpretar datos:* Este proceso constituye el centro del razonamiento estadístico, reconociendo patrones y tendencias en el conjunto de datos, y el planteamiento de inferencias y predicciones. Se pueden caracterizar dos subprocesos referidos por Curcio (1987): leer entre los datos y leer más allá de los datos.

## **PROPUESTA DE INSTRUCCIÓN**

La estrategia metodológica empleada para el diseño de las actividades de enseñanza es un acercamiento al Taller Constructivo Medina (1998), con el cual se pretende conseguir determinados propósitos educativos y formativos.

En este sentido, el taller constructivo parte de dos principios básicos:

La enseñanza se considera como un proceso intencional y planeado para lograr en los estudiantes esquemas mentales o construcciones necesarias que le permitan acceder al conocimiento y desarrollar su pensamiento matemático. Por esta razón el taller constructivo requiere de una planificación previa donde se planteen los logros esperados y todas las actividades de aprendizaje que debe realizar el estudiante bajo la mediación del profesor.

El papel del maestro es crear situaciones de aprendizaje apropiadas que le permitan al estudiante construir de manera autónoma nuevos conocimientos, a partir de los que ya conoce. Estas situaciones le posibilitarán al estudiante establecer los nexos a manera de puentes cognitivos que le favorecerán el aprendizaje significativo.

Así, todo taller constructivo implica, producción o generación de conocimiento, tanto en el estudiante como en el maestro, a través de actividades que el alumno realiza bajo la orientación del maestro. Los procesos de construcción de conocimiento a través del taller, cualifican la formación intelectual, forjan individuos críticos, autónomos, creativos, capaces de trabajar y producir en colectivo, tolerantes y a la vez exigentes consigo mismo y con los demás, pues siempre se tendrán presentes contenidos conceptuales, procedimentales y actitudinales.

Para lograr el éxito del taller se requiere participación activa y consciente de cada uno de los estudiantes y del estímulo y asesoría oportuna del maestro para identificar posibilidades y limitaciones. El papel del maestro será el de *mediador* entre las estructuras conceptuales del área y las estructuras cognitivas del alumno (Ontoria, 1994, citado por Medina (1998). En esta estrategia metodológica se requiere ante todo de un maestro dinámico y reflexivo, su papel debe ser el de *dinamizador, mediador, orientador y guía* del proceso de aprendizaje. El proceso de aprendizaje debe ser *interactivo y participativo*. El profesor dispone las condiciones y facilita las herramientas para que el alumno construya los conceptos.

La dinámica de trabajo en el taller se organiza en las siguientes etapas:

- **Revisión de conceptos previos:** Se proponen actividades para rescatar los conceptos, preconcepciones o pre teorías acerca del tema. Se revisa que los alumnos posean los conceptos preliminares, los cuales servirán para construir de forma autónoma los nuevos conocimientos.
- **Acción + Reflexión = Construcción lógica:** Trabajo individual. Toda actividad debe conducir a una reflexión. La acción está conformada por todas las actividades que propone el maestro y realiza el estudiante en forma individual, para reflexionar sobre ellas y poder descubrir regularidades que lo conducirán a la construcción de un nuevo conocimiento. Mediante preguntas que susciten nuevas preguntas el estudiante descubre características, propiedades, generalizaciones, etc. de los objetos.
- **Formulación:** Después de que el alumno construye sus propios conceptos, se busca la forma de exteriorizarlos o expresarlos. En esta etapa se emiten enunciados que pueden ser conclusiones, conjeturas, hipótesis, generalizaciones, o representaciones gráficas y simbólicas. En esta etapa, no se espera que los conceptos elaborados por



los alumnos, sean los correctos o los que maneja el profesor. Se debe apreciar y valorar la producción personal y orientar si es necesario.

- **Validación:** Confrontación en pequeños grupos o en plenaria de los procesos y resultados obtenidos en la etapa anterior. Es la oportunidad para que el estudiante aprenda a argumentar y sustentar, como también a escuchar, criticar, contra argumentar. De esta forma, desarrolla competencias comunicativas y de razonamiento lógico. Después del debate en grupo, se llega a un consenso o producción colectiva, el cual se presentará en plenaria.
- **Formalización:** El docente precisa nociones, conceptos, generalizaciones, conclusiones, lenguaje matemático, entre otras, y ofrece documentación o fuentes bibliográficas para contrastar.
- **Aplica lo aprendido:** finalmente se proponen situaciones donde el estudiante pueda exteriorizar lo aprendido y aplicarlo a situaciones del contexto.

### 3. ENFOQUE METODOLÓGICO

Para el desarrollo de este proyecto se plantea una metodología, en la que se propone un conjunto de actividades organizadas en tres fases, a través de las cuales se evidencia la manera cómo se realizó el estudio.

#### **FASE I. ANÁLISIS PRELIMINARES Y PROPUESTA DE INSTRUCCIÓN**

Este trabajo además de continuar con lo planteado por Palacios y Zambrano (2009), incorporó el diseño de una secuencia de tareas, que fue elaborado siguiendo una aproximación a los lineamientos de análisis didáctico sugeridos por Gómez (2002), quien recomienda tener en cuenta tres aspectos del los análisis previos para concretar un secuencia de a tareas para la enseñanza. En sentido se abordó un análisis desde las dimensiones del contenido, del aprendizaje y de la enseñanza. Desde la dimensión del contenido se trata de identificar y organizar los conceptos y procesos subyacentes a la conceptualización de un diagrama de caja, así como sus elementos de análisis, es decir: Tamaño de la muestra, Escala, localización, intervalos de concentración, dispersión, datos típicos extremos y datos atípicos, ya mencionados en el marco conceptual. Desde la dimensión del aprendizaje se tienen en cuenta los niveles de comprensión del grafico de caja que elaboraron Palacios y Zambrano (2009) así como los principales errores o dificultades reportados por ellos. Finalmente, desde la dimensión de la enseñanza, se hace una revisión de la propuesta de enseñanza bajo el modelo de Taller Constructivo propuesto por Medina (1998) con el fin de elaborar una secuencia de tareas para abordar la enseñanza de los gráficos de caja. Acción que responde a la dificultad de encontrar propuestas concretas que dieran cuenta de maneras de enseñar los gráficos de caja.

#### **Decisiones acerca del gráfico de caja**

Acerca de los gráficos de caja se encontraron tres versiones desde la más elaborada hasta la más básica:

Una primera versión con los límites de las zonas de anomalías ubicados a 3 recorridos intercuartiles; una segunda versión con los límites de las zonas de anomalías ubicados a 1.5 recorridos intercuartiles y una tercera que no muestra zonas de anomalías, por tanto no tiene en cuenta los datos atípicos.

Teniendo presente que en este trabajo se pretende valorar la identificación de datos atípicos por parte de los estudiantes, la primera opción es compleja para los estudiantes que no han tenido acercamiento al concepto de gráficos de cajas y la tercera no permite trabajar el concepto de datos atípicos al ignorar los límites de las zonas de anomalías. Por esta razón el tipo de gráfico de cajas que se implementó en esta propuesta es el contemplado en la segunda opción.

### **Categorías de análisis planteadas por Palacios y Zambrano**

Del proyecto de Maestría de Palacios y Zambrano (2009) se plantean las siguientes categorías de análisis del razonamiento estadístico de los estudiantes al analizar conjuntos de datos representados mediante gráficos de cajas

Estas categorías son los parámetros que permiten indagar el nivel de razonamiento estadístico por parte de los estudiantes:

- I. Tamaño de la muestra
- II. Escala:
- III. Localización
- IV. Dispersión
- V. Intervalos de concentración
- VI. Datos típicos extremos
- VII. Datos atípicos

### **Propuesta de instrucción**

Luego de la búsqueda de la literatura acerca de lo que es un gráfico de caja y de cómo se podría valorar el razonamiento estadístico de los estudiantes, se observó que no se encuentran actividades organizadas para la enseñanza del gráfico de caja; situación que requirió del diseño de una secuencia de actividades para abordar la construcción e interpretación de dichos gráficos.

Este proceso se construyó de manera conjunta entre el asesor y los autores del proyecto, estableciendo los conceptos necesarios y su secuencia para abordar el objeto de estudio, con revisiones continuas de las actividades de enseñanza realimentadas por el asesor en cada una de las tutorías, cuyo producto final se materializó a través del planteamiento, adecuación e implementación de actividades de enseñanza - aprendizaje que favorezcan en los estudiantes la comprensión e interpretación de los gráficos de caja y cada uno de

los elementos que la conforman, mediante cuatro guías con la siguiente estructura: Revisión de conceptos previos, situación problema, socialización, formalización y aplica lo aprendido, realizando su implementación durante el mes de Julio para ambos grados en sus respectivas instituciones educativas.

## **FASE II. IMPLEMENTACIÓN Y RECOLECCIÓN DE INFORMACIÓN**

Después de llevar a cabo las actividades propuestas de aprendizaje se valoró el razonamiento estadístico de los estudiantes involucrados en el estudio, mediante la aplicación del cuestionario evaluativo diseñado por Palacios y Zambrano (2009). Para el desarrollo de esta fase se tuvieron en cuenta los siguientes elementos:

### **Muestra de estudio**

Se contó con una población de 91 estudiantes de básica secundaria y media de estrato socioeconómico medio y bajo, provenientes de dos Instituciones Educativas del sector oficial, distribuidos de la siguiente manera:

Treinta y cinco estudiantes de grado noveno (903) de la Institución Educativa Departamental “Mariano Ospina Rodríguez” del municipio de Guasca C/marca, clasificada por el ICFES para el año 2011 en las pruebas Saber 11º en nivel alto y cincuenta y seis Estudiantes de grado undécimo (30 estudiantes de 1101 y 26 estudiantes de 1102) de la Institución Educativa Departamental “General Carlos Alban” del municipio de Alban C/marca, clasificada por el ICFES para el año 2011 en las pruebas Saber 11º en nivel medio.

Aunque son grupos de diferente nivel escolar, es importante resaltar que en cuanto a los conceptos estadísticos que ellos manejan son muy parecidos pues hasta ahora tienen un acercamiento a esta temática.

### **Implementación del instrumento de evaluación**

La aplicación del instrumento se realizó con los estudiantes de grado undécimo los días 9 y 10 de agosto de 2012 en los grados 1101 y 1102 respectivamente. Con los estudiantes de grado Noveno se realizó el día 9 de agosto, con un tiempo de dos horas para contestarlo.

### **FASE III ANÁLISIS DE RESULTADOS**

Una vez aplicado el instrumento de evaluación, se procede a realizar un análisis de los ítems, caracterizando el razonamiento y pensamiento estadístico para cada uno de los estudiantes de acuerdo a los niveles previamente establecidos del modelo taxonómico SOLO, realizando una revisión y ajuste de las categorías de análisis, sistematizando la información recolectada, realizando comparaciones entre las diferentes poblaciones de estudio y por último analizando las dificultades que los estudiantes presentaron.

#### **Revisión y ajuste de las categorías**

Para categorizar el razonamiento de los estudiantes se utilizó el modelo propuesto por Palacios y Zambrano (2009). Inspirado en la taxonomía SOLO En el que se relacionan los diferentes elementos y niveles de razonamiento estadístico, organizados progresivamente, empezando con el nivel más bajo y terminando con el de mayor complejidad. Al clasificar cada una de las respuestas de los estudiantes se evidenció la dificultad para ubicarlas en un nivel, por lo tanto surgió la necesidad de modificar algunos criterios de asignación del modelo original.

#### **Categorías ajustadas (Validación de categorías)**

A partir de los instrumentos aplicados a los estudiantes se realiza de manera conjunta entre el asesor y los autores del proyecto, la validación de las categorías identificando respuestas tipo por categoría, nivel de razonamiento y grado, este ejercicio permitió comparar los resultados obtenidos en contraste con las categorías ajustadas, verificando la coherencia de los criterios de asignación de las respuestas a estos niveles.

#### **Sistematización y análisis de resultados.**

Para cada uno de los elementos del razonamiento estadístico que se encuentran involucrados en las preguntas del cuestionario evaluativo, se tabuló la información mediante tablas de frecuencias con el fin de establecer el porcentaje de estudiantes categorizados en cada uno de los niveles.

#### **Comparación de poblaciones**

A partir de la información generada en la tabulación se realizaron gráficas comparativas de los resultados obtenidos por los estudiantes de los grados undécimo de la IED General Carlos Albán, noveno de la IED Técnico Comercial Mariano Ospina y el grupo de

estudiantes analizados por Palacios y Zambrano (2009), con el objetivo de validar los resultados mostrados.

## 4. RESULTADOS

### CLASIFICACIÓN DE LAS RESPUESTAS POR ÍTEMS

El instrumento de evaluación consiste en un cuestionario compuesto por diez preguntas las cuales permiten determinar el nivel de razonamiento estadístico para uno o varios elementos de análisis del gráfico de cajas, en la siguiente tabla se relacionan cada uno de los ítems con sus respectivos elementos de razonamiento:

| ELEMENTO DE RAZONAMIENTO |                             | ÍTEMS |   |   |   |   |   |   |   |   |    |
|--------------------------|-----------------------------|-------|---|---|---|---|---|---|---|---|----|
|                          |                             | 1     | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| I                        | TAMAÑO DE LA MUESTRA        | X     |   |   |   |   |   |   |   |   |    |
| II                       | ESCALA                      |       | X | X | X |   | X |   | X |   |    |
| III                      | LOCALIZACIÓN                |       | X | X |   |   |   | X |   |   | X  |
| IV                       | DISPERSIÓN                  |       |   |   |   |   |   | X |   |   | X  |
| V                        | INTERVALOS DE CONCENTRACIÓN |       | X |   |   |   | X | X |   |   |    |
| VI                       | DATOS TÍPICOS EXTREMOS      |       |   |   |   |   | X |   | X | X |    |
| VII                      | DATOS ATÍPICOS              |       |   |   |   | X |   |   |   | X |    |

Tabla 1. Categorías de razonamiento estadístico vs ítems del cuestionario Palacios y Zambrano (2009)

A continuación se presentan ejemplos representativos de cada una de las respuestas dadas por los estudiantes de los grados noveno y undécimo, ubicadas en cada uno de los niveles de cada ítem.

#### Ítem1

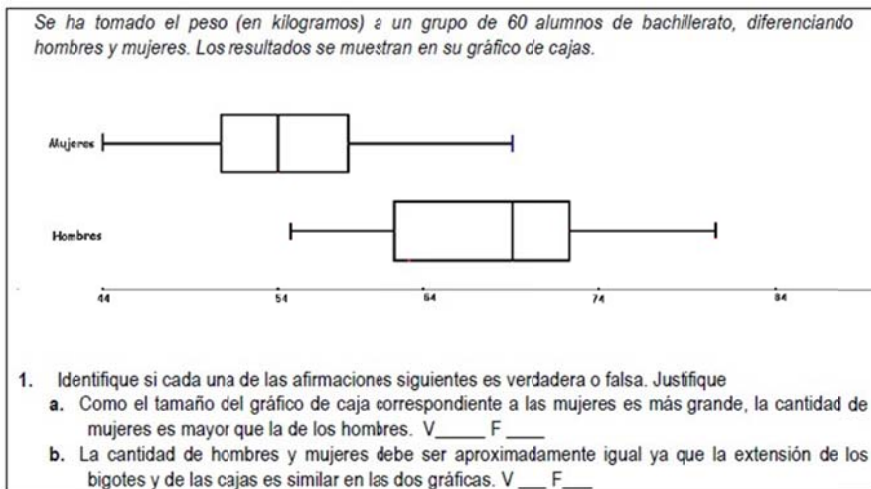


Gráfico 4. Situación uno del instrumento de evaluación

#### Ejemplo NIVEL 1

## Grado Undécimo

1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V \_\_\_ F X

Porque el gráfico de caja correspondiente a los hombres es más grande

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V \_\_\_ F X

Porque en los gráficos se puede observar que el gráfico de los hombres es mayor

## Grado Noveno

1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V X F \_\_\_

porque hay estatuas de mujeres muy segidas X

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V X F \_\_\_

porque la extensión de los bigotes es similar. X

## Ejemplo NIVEL 2

### Grado Noveno

1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V \_\_\_ F X

Porque no se puede saber cuántas mujeres son X

## Ejemplo NIVEL 3

### Grado Undécimo

1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V \_\_\_ F X

no, porque esta diferencia de tamaño es debido a que está concentrada y la de los hombres es más dispersa y además estas gráficas nos ayudan a calcular una especie de promedios, y se nos dificulta calcular la cantidad de hombres y la de mujeres

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V \_\_\_ F X

porque los bigotes nos señalan el dato mínimo típico y el dato máximo típico

### Grado Noveno



1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V    F X  
 por que si el grafico es mas amplio es porque los resultados están más dispersos.

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V    F X  
 porque las cajas representan los resultados dispersos y acumulados y no quiere decir que que hayan igual numero de personas.

Ejemplo NIVEL 4  
 Grado Undécimo

1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V    F X  
 porque el grafico de cajas nos indica cual es la dispersion de los datos en cambio si queremos saber la cantidad de mujeres encuestadas acudimos al diagrama de barras

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V    F X  
 por que los bigotes no indican el limite inferior y superior

Ítem 2

2. ¿Cuál grupo presenta mayores pesos? Mujeres    Hombres    Justifique

Ejemplo NIVEL 1  
 Grado Noveno

2. ¿Cuál grupo presenta mayores pesos? Mujeres X Hombres    Justifique  
 por que desde la mediana las mujeres pesan de 70 kilogramos.

Ejemplo NIVEL 2  
 Grado Undécimo

2. ¿Cuál grupo presenta mayores pesos? Mujeres    Hombres X Justifique  
 porque la caja de los hombres presenta un número de datos con más valor

2. ¿Cuál grupo presenta mayores pesos? Mujeres    Hombres X Justifique  
 por que el dato minimo esta muy por encima del dato minimo en las mujeres e igualate al dato max

Grado Noveno

2. ¿Cuál grupo presenta mayores pesos? Mujeres    Hombres X Justifique  
 porque las mujeres pesan 69 kg y los hombres 80 kg

Ejemplo NIVEL 3  
Grado Undécimo

2. ¿Cuál grupo presenta mayores pesos? Mujeres \_\_\_\_\_ Hombres X Justifique  
Porque en el gráfico está más a la derecha el de los hombres que el de las mujeres.

Ejemplo NIVEL 4  
Grado Noveno

2. ¿Cuál grupo presenta mayores pesos? Mujeres \_\_\_\_\_ Hombres X Justifique  
porque el de los h. está entre 54 y 81 kg y el de los M está entre 44 y 64 kg.

Ítem 3

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:  
a. ¿A la persona que pesa 54 Kg?  
b. ¿A la persona que pesa 70 Kg?

Ejemplo NIVEL 1  
Grado Undécimo

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:  
a. ¿A la persona que pesa 54 Kg?  
Es un peso que está entre el rango como de hombres y mujeres.

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:  
a. ¿A la persona que pesa 54 Kg?  
que esta persona es del grupo de mujeres de donde está distribuido y que pesa menos que hombre.

b. ¿A la persona que pesa 70 Kg?  
es del grupo de hombres y que es el peso más que la mujer.

Grado Noveno

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:  
a. ¿A la persona que pesa 54 Kg?  
que es el peso acumulado en las mujeres. ?

b. ¿A la persona que pesa 70 Kg?  
que es el peso acumulado en los hombres. ?

Ejemplo NIVEL 2  
Grado Undécimo

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

Es la persona que pesa menos en los hombres y en las mujeres es el Q<sub>2</sub>.

b. ¿A la persona que pesa 70 Kg?

Es la persona más pesada y en los hombres es el Q<sub>2</sub>.

#### Grado Noveno

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

Se trata de una mujer de acuerdo a su peso coincide con la mediana de las mujeres.

b. ¿A la persona que pesa 70 Kg?

Se trata de un hombre de acuerdo a su peso coincide con la mediana.

#### Ejemplo NIVEL 3

#### Grado Noveno

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

En las mujeres es la mediana y en los hombres es un dato atípico.

b. ¿A la persona que pesa 70 Kg?

En las mujeres es el límite superior y en los hombres es la mediana.

#### Nivel 4

#### Grado Undécimo

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

que es el Q<sub>2</sub> en las mujeres, se puede decir que el promedio en los hombres este dato no está presente siendo el hombre menos pesado del grupo.

b. ¿A la persona que pesa 70 Kg?

en los hombres se podría decir que es la mediana se encuentra como un dato normal en las mujeres que este es el dato máximo típico.

#### Grado Noveno

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

que es la me o Q<sub>2</sub> de las mujeres mientras que en los hombres empieza en 54.

b. ¿A la persona que pesa 70 Kg?

que en los hombres es la me o Q<sub>2</sub> y en las mujeres termina en 70.

#### Ítem 4

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.
- b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

Ejemplo NIVEL 1  
Grado Undécimo

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

aproximadamente el 90% de las personas pesa más de 50 kg

- b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

aproximadamente el 82% de las personas pesa menos de 72 kg

Grado Noveno

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

El 50% ya que es un peso normal en las mujeres y en los hombres es bajo

- b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

El 70% de los estudiantes ya que después es un peso normal

Ejemplo NIVEL 2  
Grado Undécimo

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

En los hombres el % es de 100% porque todos los hombres pesan mayor a este peso. En las mujeres un 50% pesa el dato Q<sub>3</sub> esto casi ubicado en este peso.

- b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

En los hombres casi un 90%. Pese a este peso esto ubicado al Q<sub>3</sub> En las mujeres el 100% porque todos las mujeres pesan menos de 72 kg

Grado Noveno

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

75% porque es de el Q<sub>1</sub> 25% Q<sub>2</sub> = 25% Q<sub>3</sub> = 25%.  
25% + 25% + 25% = 75%

- b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

En los hombres: 75%. En las mujeres: 100%

Ejemplo NIVEL 3  
Grado Undécimo

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

Mujeres = un 75% de la población peso más del 50kg. Ya que comienza desde 0 y hombres = el 100% de la población masculina peso más de 50kg. Ya que su gráfica está con pesos más elevados.

b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

Mujeres = el 100% de la población femenina pesa menos de 72kg. hombres = calculo que un 75% de la población masculina pesa menos del 72kg.

### Grado Noveno

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

en los ~~mujeres~~ <sup>hombres</sup> el 100% peso más de 50kg y en las mujeres el 75%.

b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

en los hombres el 75% peso menos de 72kg y en las mujeres el 100%.

### Ítem 5

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

### Ejemplo NIVEL 1

#### Grado Undécimo

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

por que pueden ser también el mismo porcentaje o la misma cantidad.

#### Grado Noveno

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

No hay datos atípicos X

### Ejemplo NIVEL 2

#### Grado Undécimo

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

solo en los hombres hay un dato atípico, un dato atípico es aquel que está fuera del rango de los datos

#### Grado Noveno

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

si hay y en el grupo o los hombres.

### Ejemplo NIVEL 3

#### Grado Undécimo

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

en los pesos de las mujeres no hay datos atípicos; en los pesos de los hombres si hay un dato atípico porq' sobre pasa el peso normal.

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique por qué son atípicos  
en los hombres existe un dato atípico porque está fuera de una zona normal de los pesos de los hombres. y este es superior a 84 kg

Grado Noveno

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique por qué son atípicos  
si hay uno en los hombres que está entre 90 y 94, q' está en sobrepeso

Ítem 6

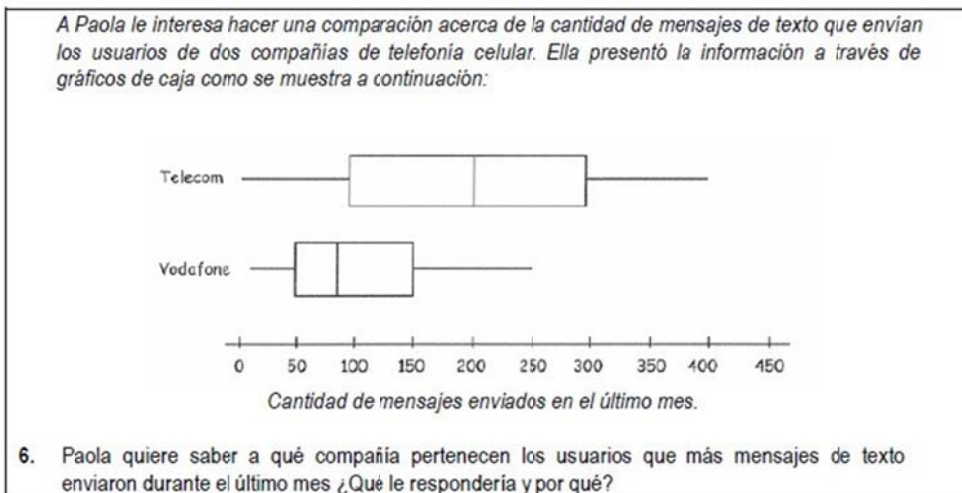


Gráfico 5. Situación dos del instrumento de evaluación

Ejemplo NIVEL 1

Grado Undécimo

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?  
telecom porque este fue el que más envió mensajes en el último mes

Grado Noveno

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?  
Vodafone por que tiene los datos concentrados y no dispersos

Ejemplo NIVEL 2

Grado Undécimo

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?  
Pienso que telecom por que a pedido obtener 400 mensajes, Vodafone solo pudo obtener 250

Grado Noveno

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?

telecom porque el es que tiene más que vodafone durante el mes

### Ejemplo NIVEL 3

Grado Noveno

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?

a telecom porque va de 0 usuarios a 400 usuarios

### Ejemplo NIVEL 4

Grado Noveno

6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?

Telecom x q" esta entre 100 y 300 mientras que Vodafone esta entre 50 y 150

### Ítem 7

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

### Ejemplo NIVEL 1

Grado Undécimo

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

200 para telecom, porque el dato esta sustituido en 92 para vodafone

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

### Grado Noveno

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

la cantidad de mensajes es de 450 en el mes. X

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

vodafone es el que tiene mayor número de mensajes que envia. X

## Ejemplo NIVEL 2

### Grado Undécimo

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

Telecom: 100 - 300, porque desde estos valores comienza la caja y cada caja vale el 26%.

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

Falso, porque los datos que se encuentran en Vodafone están más concentrados que que el tamaño del gráfico es menor.

### Grado Noveno

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

Entre 100 y 300, por que cada uno tiene un 25%.

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

es falso la que tiene mas dispersion es Telecom.

## Ejemplo NIVEL 3



## Grado Noveno

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

Telecom: el 50% está entre 93 y 300, o sea 200.

Vodafone: el 50% está entre 50 y 150, o sea 85.

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

No porque la caja de Telecom es más grande o sea que envían más usuarios mensajes que la de los de Vodafone.

## Ejemplo NIVEL 4

### Grado Noveno

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

Telecom = 100 y 300 es la caja y no los bigotes más dispersa

Vodafone = 50 y 150 porque está más acumulada

b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

No Vodafone tiene mayor agrupación y Telecom está más dispersa que Vodafone.

## Ítem 8

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

## Ejemplo NIVEL 1

### Grado Undécimo

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

En la compañía Telecom el usuario que más mensajes envió fue 300 mjs en el último mes.  
En la compañía Vodafone el usuario que más mensajes envió fue 150 mjs en el último mes.

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

En la compañía Telecom el usuario que menor consumo presentó fue de 100 mjs.  
En la compañía Vodafone

el usuario que menor consumo presentó fue de 50 mjs.

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

Que la compañía Telecom es la compañía en la que más cantidad de mensajes de texto se enviaron.

## Grado Noveno

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

Telecom: 300 mensajes R  
vodafone: 150 mensajes

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

Telecom: 100 mensajes x Cero  
vodafone: 50 mensajes x

### Ejemplo NIVEL 2

#### Grado Undécimo

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta

Telecom 400 vodafone 250 ya que es el límite o dato máximo

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

Telecom 0 vodafone 0 ya que son los límites mínimos

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

Que ambas compañías son datos mínimos o cero pero del que más se consumen mensajes es de telecom ya que su dato máximo es mayor al de vodafone

#### Grado Noveno

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

Telecom: 400 ✓  
vodafone: 250 ✓

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

Telecom: en los bigetes 0 ✓  
vodafone: en los bigetes 0 ✓

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

que telecom es el que más mensajes se enviaron en el último mes que telecom es el que más mensajes se enviaron en el último mes

### Ejemplo NIVEL 3

#### Grado Noveno

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

el usuario que envió más mensajes a telecom fue de 400 y de vodafone fue 250 ya que es donde terminan los diagramas de caja

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

en las 2 compañías el que tuvo menor fue 0 ✓

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

que el de telecom fue más alto que el de vodafone y que el de menor las 2 compañías está iguales en 0

### Ejemplo NIVEL 4

#### Grado Undécimo

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.  
 en telecom es de 400 mensajes pues este es su dato máximo típico  
 en vodafone es de 250 pues este es su dato máximo típico.
- b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.  
 las dos compañías presentan como dato mínimo cero
- c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?  
 que la empresa telecom debe ser la más prospera pues el número de mensajes es muy alto y casi duplica a vodafone

### Grado Noveno

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.  
 Telecom = 400 que es el límite superior para esta compañía  
 Vodafone = 250 que es el límite superior para esta compañía
- b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.  
 Tanto para telecom como para vodafone es de 0 usuarios o sea límite inferior
- c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?  
 pues que los usuarios de telecom y Vodafone es el límite inferior, 0

### Ítem 9

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta
- b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

### Ejemplo NIVEL 1

#### Grado Undécimo

El \_\_\_% de los estudiantes no contestaron este ítem

- no es un dato atípico porque no está dentro del parámetro de los datos
- b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.  
 no necesariamente y porque los gráficos son totalmente diferentes por eso no afecta y porque no está dentro del parámetro de los datos.

#### Grado Noveno

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta  
 Si ya es el último mensaje que envió X
- b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.  
 No

Ejemplo NIVEL 2  
Grado Undécimo

SE PUEDE UN DATO ATÍPICO YA QUE DEBERÍA SER LA DISTRIBUCIÓN DE LOS DATOS EN LAS EMPRESAS

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

NO, POR QUE SE PONDRÍA UN PUNTO ESPECÍFICO INDICANDO QUE ESTE ES UN DATO ATÍPICO

Grado Noveno

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta

Si sería un atípico por que se saldría de rango ya el tiene

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

si afectaría su promedio por que aumentaría por ser mayor q.

Ejemplo NIVEL 3  
Grado Undécimo

Si, ya que no está dentro del rango de los datos que aparecen comúnmente allí, sino que es un caso especial

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

no en mucho ya que por ser un dato atípico solo se pondría un punto allí, pero la caja no la afectaría

Grado Noveno

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta

si es un dato atípico porque 450 sobresale de los bigotes que corresponde que 450 es un dato atípico.

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

si porque estaría más disperso y obtendría una mayor cantidad de consumo.

Ejemplo NIVEL 4  
Grado Noveno

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta

Si porque estaría fuera del límite superior de esta línea telefónica.

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

Si porque se alargaría o se expandiría más el bigote

Ítem 10

a. ¿En qué compañía los valores son más cercanos entre sí? \_\_\_\_\_

b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? \_\_\_\_\_

Ejemplo NIVEL 1

### Grado Undécimo

10. a. ¿En qué compañía los valores son más cercanos entre sí? 50 y 50.

b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía?

Que a pesar de todo no importa la cantidad de mensajes que envíen por la 2 compañía va a ser el mismo porcentaje.

### Grado Noveno

10. a. ¿En qué compañía los valores son más cercanos entre sí? 50 y 50.

b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? que telecom esta igual

### Ejemplo NIVEL 2

#### Grado Undécimo

### Ejemplo NIVEL 3

#### Grado Undécimo

10. a. ¿En qué compañía los valores son más cercanos entre sí? vodafone.  
b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? Del 50% inferior que son datos agrupados y el 50% superior que son datos más dispersos.

10. a. ¿En qué compañía los valores son más cercanos entre sí? en vodafone.  
b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? que el 50% superior esta mas disperso que el inferior.

### Grado Noveno

10. a. ¿En qué compañía los valores son más cercanos entre sí? Vodafone  
b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? el 50% inferior estaria mas agrupado y el 50% superior estaria mas disperso.

## TIPOS DE RESPUESTAS DE LOS ESTUDIANTES POR CATEGORÍAS

A continuación se presentan las categorías de razonamiento estadístico al comparar dos conjuntos de datos representados por gráficos de caja planteadas por Palacios y Zambrano (2009) ; las cuales permiten ubicar el nivel de las respuestas obtenidas por los estudiantes (Ver Tabla 2).

Para valorar el nivel de razonamiento de los estudiantes se presentó la dificultad de ubicar las respuestas obtenidas en los niveles originalmente planteados. A partir de las respuestas obtenidas por los estudiantes se observó que los criterios para establecer el

nivel de en cada una de las categorías no se ajustaba a la información obtenida de los ítems del cuestionario; razón por la cual se decidió adecuar la matriz (Ver Tabla 3).

Con el fin de verificar la consistencia de la nueva matriz se decidió confrontarla con ejemplos de respuestas obtenidas tanto por los estudiantes de noveno como de undécimo grado para validar los cambios realizados en la misma, (Ver Tabla 4).La única categoría que no sufrió cambios fue la correspondiente a Intervalos de Concentración.

**CATEGORÍAS DE RAZONAMIENTO ESTADÍSTICO AL COMPARAR DOS CONJUNTOS DE DATOS REPRESENTADOS POR GRÁFICOS DE CAJA  
PLANTEADA POR PALACIOS Y ZAMBRANO 2009**

|                                    | <b>NIVEL 1<br/>IDIOSINCRÁSICO<br/>Preestructural</b>   | <b>NIVEL 2<br/>TRANSICIONAL<br/>Uniestructural</b>   | <b>NIVEL 3<br/>CUANTITATIVO<br/>Multiestructural</b>  | <b>NIVEL 4<br/>ANALÍTICO<br/>Relacional</b>  |
|------------------------------------|--|--|---|--|
| <b>TAMAÑO DE LA MUESTRA</b>        | Afirma que el tamaño de la muestra está relacionado con la amplitud y forma del gráfico.         | Independiza el tamaño de la muestra de la interpretación inicial del gráfico, asumiendo un solo gráfico de caja.                   | Caracteriza el efecto del muestreo en dos gráficos de caja, pero no establece relaciones entre ellos.   | Reconoce que no se puede identificar el tamaño de la muestra en una lectura del gráfico, pero relaciona el tamaño de los gráficos para obtener conclusiones.   |
| <b>ESCALA</b>                      | No relaciona la escala de la recta numérica con la representación gráfica de cada caja.          | Identifica los valores de los cinco resúmenes en un solo gráfico de cajas  | Reconoce e interpreta separadamente para cada gráfico cada uno de las medidas de posición, pero no compara dichos resúmenes entre los dos gráficos.           | Relaciona los valores de las medidas de resumen de cada gráfico y establece conclusiones acerca de las características de ambas distribuciones de datos.   |
| <b>LOCALIZACIÓN</b>                | No identifica la mediana, en el gráfico de caja, como centro porcentual de un conjunto de datos. | Reconoce la mediana como una medida de tendencia central representativa de todo el conjunto de datos en un solo conjunto de datos. | Describe las medianas de dos conjuntos de datos, pero no establece relaciones entre ellos.  | Establece relaciones entre dos gráficos de caja al comparar sus dos medianas.  |
| <b>DISPERSIÓN</b>                  | No identifica el rango intercuartílico en un gráfico de caja.                                    | Identifica el rango intercuartílico como una medida de dispersión en un solo gráfico de caja.                                      | Reconoce los rangos intercuartílicos en dos gráficos de caja, pero no establece relaciones entre estas distribuciones.  | Identifica la dispersión como una característica de la forma del conjunto de datos relacionando los rangos intercuartílicos de los gráficos de caja.   |
| <b>INTERVALOS DE CONCENTRACIÓN</b> | No reconoce las agrupaciones de los datos respecto al recorrido de los valores de la variable.   | Describe el comportamiento de un conjunto de datos en particiones específicas de un solo gráfico de caja.                          | Identifica el comportamiento de la agrupación de los datos en dos gráficos de caja, pero sin establecer relaciones entre ellos.                               | Compara la concentración de los datos en zonas específicas de dos conjuntos de datos y establece conclusiones acerca de la relación entre ambas distribuciones de datos.                                 |
| <b>DATOS ADYACENTES</b>            | No identifica los valores que son límites de las cercas.   | Identifica el mayor y el menor dato considerados en los segmentos exteriores de un gráfico de caja.                                | Describe los valores considerados como límites de los bigotes en dos conjuntos de datos por separado, sin establecer relaciones entre los conjuntos de datos. | Establece conclusiones acerca de la distribución de dos conjuntos de datos al comparar los datos adyacentes de ambas distribuciones.   |
| <b>DATOS ATÍPICOS</b>              | No tiene en cuenta la presencia de datos atípicos en la distribución de datos.                   | Considera y caracteriza los datos atípicos en un solo gráfico de caja.   | Determina los datos atípicos en los gráficos de caja para ambas distribuciones de datos pero sin establecer relaciones.                                       | Caracteriza el efecto que tiene la inclusión de los valores atípicos en la distribución del conjunto de datos y establece conclusiones al comparar los cambios en las distribuciones por esta inclusión. |

Tabla 2. Categorías de razonamiento estadístico Palacios y Zambrano (2009)

**CATEGORÍAS DE RAZONAMIENTO ESTADÍSTICO AL COMPARAR DOS CONJUNTOS DE DATOS REPRESENTADOS POR GRÁFICOS DE CAJA MODIFICADA**

|                                    | <b>NIVEL 1<br/>IDIOSINCRÁSICO<br/>Preestructural</b>   | <b>NIVEL 2<br/>TRANSICIONAL<br/>Uniestructural</b>  | <b>NIVEL 3<br/>CUANTITATIVO<br/>Multiestructural</b>  | <b>NIVEL 4<br/>ANALÍTICO<br/>Relacional</b>  |
|------------------------------------|--|---|---|--|
| <b>TAMAÑO DE LA MUESTRA</b>        | Afirma que el tamaño de la muestra está relacionado con la amplitud y forma del gráfico.       | Reconoce que el tamaño de la muestra no se puede establecer a partir del gráfico de caja. Argumentando con un solo elemento o característica del gráfico de caja. | Afirma que el tamaño de la muestra no se puede establecer a partir del gráfico de caja y justifica con más de un elemento o característica del gráfico de caja. | Reconoce que no se puede identificar el tamaño de la muestra en la lectura del gráfico de caja, argumentando esta afirmación con más de un elemento o característica del gráfico de caja y propone otra manera de hallar el tamaño de muestra. |
| <b>ESCALA</b>                      | No relaciona la escala de la recta numérica con la representación gráfica de cada caja.        | Relaciona la escala de la recta numérica con alguno de los valores (Min,Q1,Q2,Q3,Max,Datos Atípicos) de un SOLO gráfico de caja                                   | Relaciona la escala de la recta numérica con más de un valor (Min,Q1,Q2,Q3,Max,Datos Atípicos) de un SOLO gráfico de caja                                       | Relaciona los valores de las medidas de resumen de cada gráfico y establece conclusiones acerca de las características de ambas distribuciones de datos.   |
| <b>LOCALIZACIÓN</b>                | No ubica ninguna medida de localización en un grafico de cajas                                 | Ubica al menos una medida de localización en los gráficos de cajas  | Ubica dos o más medidas de localización en los gráficos de cajas  | Relaciona dos o más medidas de localización en los gráficos de cajas   |
| <b>DISPERSIÓN</b>                  | No relaciona la dispersión de los datos con la amplitud del grafico de caja                    | Relaciona la dispersión de los datos con la amplitud del grafico de caja  | Determina la dispersión de los datos en los diferentes intervalos que se pueden formar con los cinco resúmenes de un grafico de caja                            | Establece conclusiones a partir del análisis de la dispersión de los datos obtenida con los intervalos que se pueden formar con los cinco resúmenes de un grafico de caja  |
| <b>INTERVALOS DE CONCENTRACIÓN</b> | No reconoce las agrupaciones de los datos respecto al recorrido de los valores de la variable. | Describe el comportamiento de un conjunto de datos en particiones específicas de un solo gráfico de caja.   | Identifica el comportamiento de la agrupación de los datos en dos gráficos de caja, pero sin establecer relaciones entre ellos.                                 | Compara la concentración de los datos en zonas específicas de dos conjuntos de datos y establece conclusiones acerca de la relación entre ambas distribuciones de datos.   |



|                                   |   |   |   |  |
|-----------------------------------|---|---|---|--|
| <b>DATOS TIPICOS<br/>EXTREMOS</b> | No identifica los límites de los bigotes.   | Identifica el mayor y el menor dato considerados en los segmentos exteriores de un gráfico de caja. | Describe los valores considerados como límites de los bigotes en dos conjuntos de datos por separado, sin establecer relaciones entre los conjuntos de datos. | Establece conclusiones acerca de la distribución de dos conjuntos de datos al comparar los datos adyacentes de ambas distribuciones.   |
| <b>DATOS ATÍPICOS</b>             | No ubica ni tiene en cuenta la presencia de valores atípicos en la distribución de datos. | Ubica y tiene en cuenta la presencia de valores atípicos en la distribución de datos.               | Interpreta y Justifica la presencia de valores atípicos en una distribución de datos  | Caracteriza el efecto que tiene la inclusión de los valores atípicos en la distribución del conjunto de datos y establece conclusiones al comparar los cambios en las distribuciones por esta inclusión. |

Tabla 3. Categorías de razonamiento estadístico modificada

**RESPUESTAS OBTENIDAS POR LOS ESTUDIANTES POR CATEGORIA Y NIVEL**

|                             | <b>PREGUNTA</b>   | <b>Grado</b>    | <b>NIVEL 1<br/>IDIOSINCRÁSICO<br/>Preestructural</b>                                | <b>NIVEL 2<br/>TRANSICIONAL<br/>Uniestructural</b>  | <b>NIVEL 3<br/>CUANTITATIVO<br/>Multiestructural</b>  | <b>NIVEL 4<br/>ANALÍTICO<br/>Relacional</b>  |
|-----------------------------|---|-----------------|---|---|---|--|
| <b>TAMAÑO DE LA MUESTRA</b> | La cantidad de mujeres es mayor que la de los hombres?                              | <b>UNDECIMO</b> | Falso<br><br>Porque el grafico de cajas correspondiente a los hombres es más grande | Falso<br><br>Son igualmente proporcionales sino que los hombres poseen más peso                 | Falso<br><br>No porque esta diferencia de tamaños es debida a que está concentrada y la de los hombres es más dispersa y además estas gráficas nos ayudan a calcular los promedios y se nos dificulta calcular la cantidad de hombres y mujeres | Falso<br><br>Porque el grafico de cajas nos indica cual es la dispersión de los datos; en cambio si queremos saber la cantidad de mujeres encuestadas acudimos al diagrama de barras |
|                             |   | <b>NOVENO</b>   | Verdadero<br><br>Porque hay estaturas de mujeres muy seguidas                       | Falso<br><br>Porque no se pueden saber cuántas mujeres son                                      | Falso<br><br>Porque si el gráfico es más amplio es porque los resultados están más dispersos  |  |
| <b>ESCALA</b>               | ¿Cuál grupo presenta mayores pesos?<br>Mujeres _____<br>Hombres _____<br>Justifique | <b>UNDECIMO</b> | Hombres<br><br>Porque el grafico muestra que hay mayor cantidad de pesos            | Hombres<br><br>Porque la caja de los hombres presenta un mayor número de datos con mayor valor. | Hombres<br><br>Porque el gráfico que está más a la derecha es el de los hombres que el de las mujeres.  | Hombres<br><br>Porque su dato mínimo está muy por encima de las mujeres; e igualmente su dato máximo   |
|                             |   | <b>NOVENO</b>   | Mujeres<br><br>Porque desde la Mediana las mujeres pesan 70 kilos                   | Hombres<br><br>Porque las mujeres pesan 69 kg y los hombres 80 kg                               |   | Hombres<br><br>Porque el de los hombres está entre 54 y 81 kg y el de las mujeres está entre 44 y 64 kg  |

|                     |  |   |   |   |  |  |
|---------------------|--|---|---|---|--|--|
| <b>LOCALIZACIÓN</b> | ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:<br><br>a. ¿A la persona que pesa 54 Kg?<br><br>b. ¿A la persona que pesa 70 Kg?   | <b>UNDECIMO</b>   | a. Es un peso que está en el rango tanto de hombres como de mujeres   | a. Es la persona que pesa menos de los hombres y en las mujeres es el Q2  | a. Que en las mujeres es el promedio (Q2) normal y en los hombres es el peso más inferior (Límite inferior)<br><br>b. Que en los hombres es el promedio (Q2) Normal y en las mujeres que es el peso máximo al cual es normal llegar. | a. Que es el Q2 de las mujeres, se puede decir que el promedio en los hombres; ese dato no está presente; sería el hombre menos pesado del grupo<br><br>b. En los hombres se podría decir que es la mediana se encuentra como un dato normal; en las mujeres este sería el dato máximo típico. |
|                     | <b>NOVENO</b>  | a. Que es el peso acumulado de las mujeres.<br><br>b. Que es el peso acumulado de los hombres | a. Sería una mujer de acuerdo a su peso coincide con la mediana de las mujeres<br><br>b. Sería un hombre, de acuerdo a su peso coincide con la mediana                    | a. En las mujeres es la mediana y en los hombres es un dato atípico<br><br>b. En las mujeres es el límite superior y en los hombres es la mediana | a. Que es la mediana o Q2 de las mujeres mientras que en los hombres comienza en 54<br><br>b. Que en los hombres es la Me o Q2 y en las mujeres termina en 70  |  |
| <b>DISPERSIÓN</b>   | b. Paola afirma que "Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom". Confirme o refute esta afirmación usando la información presentada en los gráficos de caja. | <b>UNDECIMO</b>   | Porque Vodafone tiene menor cantidad de mensajes que Telecom porque Telecom tiene una cantidad de mensajes de igual porcentaje; en cambio Vodafone tiene mayor porcentaje | Falso; porque los datos que se encuentran en Vodafone están más concentrados ya que el tamaño del gráfico es menor                                | La Mayor dispersión del número de mensajes es mayor en Telecom ya que la caja y los bigotes son más extensos   | Falso porque la gráfica más dispersa es la de Telecom y la de Vodafone se encuentra más concentrada pues sus clientes tienen mayor similitud en el número de mensajes.   |
|                     |  | <b>NOVENO</b>   | Vodafone es el que tiene mayor número de mensajes que envía   | Es falso, la que tiene mayor dispersión es Telecom  | No, porque la caja de Telecom es más grande, o sea que envían más usuarios mensajes que las de Vodafone  | No, Vodafone tiene una mayor agrupación y Telecom está más dispersa que Vodafone   |

|                                    |  |                 |  |  |  |   |
|------------------------------------|--|-----------------|--|--|--|---|
| <b>INTERVALOS DE CONCENTRACIÓN</b> | ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué? | <b>UNDECIMO</b> | Vodafone está entre 50 y Telecom entre 100 y 200 porque los dos están comprendidos entre Q1 y Q2   | Telecom: 100-300 porque desde estos valores comienza la caja y cada caja vale el 25% | En el de Telecom se encuentra entre 100 y 300, el punto central es 200. Para Vodafone está entre 50 y 150  | No se registran respuestas en esta categoría  |
|                                    |  | <b>NOVENO</b>   | La cantidad de mensajes es de 450 en el mes.   | Entre el cuartil 1 y 3 porque cada uno tiene el 25%                                  | Telecom: El 50% está entre 93 y 200<br>Vodafone: El 50% está entre 50 y 150  | Telecom: 100 y 300 es la caja y no los bigotes más dispersos<br><br>Vodafone: 50 y 150 porque está más acumulada                                  |
| <b>DATOS TÍPICOS EXTREMOS</b>      | a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.                   | <b>UNDECIMO</b> | En la compañía Telecom el usuario que más mensajes envió fue 300 mjs en el último mes; en la compañía Vodafone el usuario que más mensajes envió fue de 150 en el último mes | Telecom 400 Vodafone 250 ya que es el límite o dato máximo                           | En Telecom 400 mensajes según los extremos del bigote de la caja superior, Vodafone 250 mensajes según el extremo del bigote de la caja superior.                  | En Telecom es de 400 mensajes pues este es su dato máximo típico en Vodafone es de 250 pues este es su dato máximo típico                         |
|                                    |  | <b>NOVENO</b>   | Telecom: 300 mensajes<br>Vodafone: 150 Mensajes  | Telecom: 400<br>Vodafone: 250  | El usuario que envió más mensajes por Telecom fue de 400 y de Vodafone fue de 250 porque es donde terminan los diagramas de cajas                                  | Telecom: 400 que es el límite superior de esta compañía<br><br>Vodafone: 250 que es el límite superior de esta compañía                           |
|                                    | b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su                        | <b>UNDECIMO</b> | En la compañía Telecom el usuario que menor consumo presentó fue de 100 mensajes; en la compañía Vodafone, el usuario que menor consumo presentó fue de 50 mensajes.         | Telecom 0 Vodafone 0 ya que son los límites mínimos                                  | Telecom 0 mensajes según el extremo del bigote de la caja inferior. Vodafone menos de 50 mensajes porque el extremo del bigote inferior no alcanza a llegar a cero | Las dos compañías presentan como dato típico cero<br><br>Que la empresa Telecom debe ser la más próspera pues el número de mensajes es muy alto y |

|                       |  |                 |  |  |   |  |
|-----------------------|--|-----------------|--|--|---|--|
|                       | respuesta.   |                 |  |  |   | casi duplica a Vodafone  |
|                       |  | <b>NOVENO</b>   | Telecom: 100 mensajes<br>Vodafone: 50 Mensajes   | Telecom: En los bigotes 0<br>Vodafone: En los bigotes 0                                  | En las dos compañías el que tuvo menor fue cero   | Tanto para Telecom como para Vodafone es de cero usuarios o sea el límite inferior.                      |
| <b>DATOS ATÍPICOS</b> | 9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta | <b>UNDECIMO</b> | No porque habría compatibilidad entre las dos empresas que tiene que ser mayor o menor | Sí porque sería el usuario que más hubiera marcado msj en un mes a través de la compañía | Sería un dato atípico porque el límite superior de esta compañía es de 250                                    | Si pues supera por gran cantidad al dato máximo típico y entraría en una zona atípica                    |
|                       |  | <b>NOVENO</b>   | Sí porque es el último mensaje que envió   | Sí sería un atípico porque se saldría del rango que el tiene                             | Sí es un dato atípico porque 450 se sobrepasa de los bigotes que hace corresponder a 450 como un dato atípico | Si porque estaría por fuera del límite superior de esta línea telefónica                                 |
|                       | b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.  | <b>UNDECIMO</b> | No porque el grafico muestra cada una de las empresas por aparte                       | Se señalaría con un punto este dato que es atípico                                       | Sí porque tendría que meter entre la gráfica el dato atípico aunque no dañaría su estructura central          | No, en mucho, ya que por ser un dato atípico solo se pondría un punto allí, pero la caja no la afectaría |
|                       |  | <b>NOVENO</b>   | No   | Sí aumentaría su promedio porque aumentaría por ser mayor que el de Telecom              | Sí porque estaría más disperso y obtendría una mayor cantidad de consumo                                      | Si porque se alargaría o se expandiría más el bigote   |

Tabla 4. Respuestas de los estudiantes por categoría y nivel.

## PRESENTACIÓN Y ANÁLISIS DE RESULTADOS

A continuación se realiza un análisis donde se confrontan los resultados obtenidos por la población de estudio de Palacios y Zambrano (2009), los estudiantes de los grados undécimo y noveno y un grupo total (undécimo y noveno) que reúne la población de interés de este proyecto. Dado que la cantidad de estudiantes en cada uno de los grupos son de diferente tamaño, se decidió presentar los resultados de manera porcentual para hacerlos igualmente comparables.

En los siguientes gráficos se indica el nivel comparativo de las poblaciones analizadas para cada una de las categorías del razonamiento estadístico.

### Con respecto al Tamaño de la Muestra

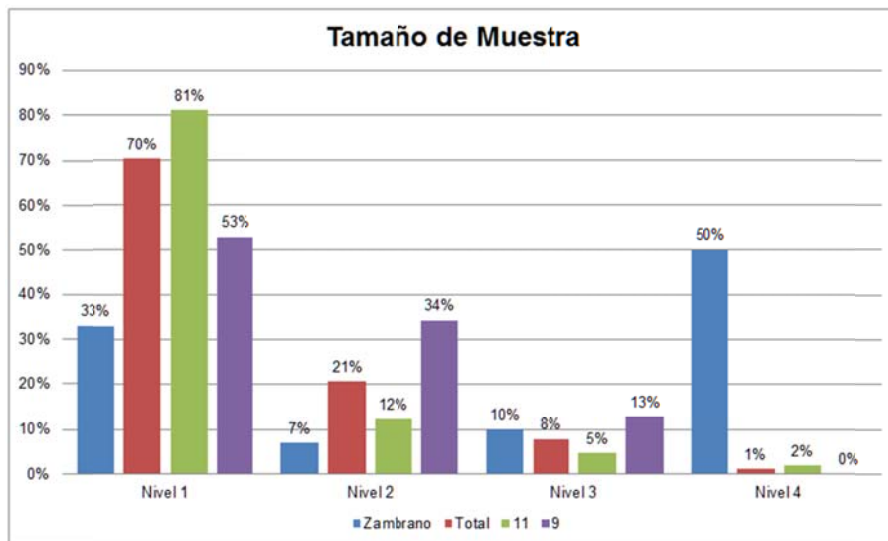


Gráfico 6. Tamaño de muestra

En esta categoría se observa que el nivel predominante es el idiosincrásico, aproximadamente el 70% de los estudiantes del grupo total se encuentran en este nivel, además es apreciable la diferencia entre los grados noveno y undécimo. En contraste con la población de estudio de Palacios y Zambrano en la cual, la mitad de los estudiantes se encuentran categorizados en el nivel relacional.

## Con respecto a la Escala

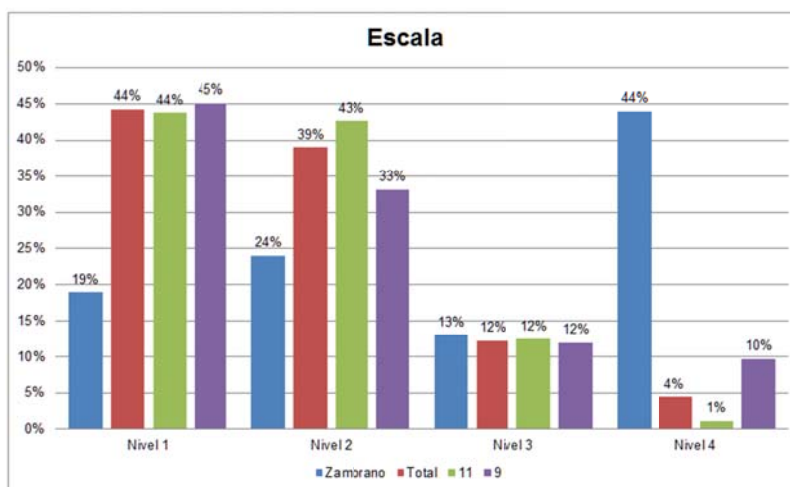


Gráfico 7. Escala

La mayoría de los estudiantes de la población del presente estudio se ubican en los niveles uno y dos, destacándose que los dos grados noveno y undécimo tienen un comportamiento similar. Además se observa una notable diferencia comparados con la población de Palacios y Zambrano (2009) pues un 44% de esta población se encuentra ubicado en el nivel relacional en contraste con un 4% de la población de estudio.

## Con respecto a la Localización

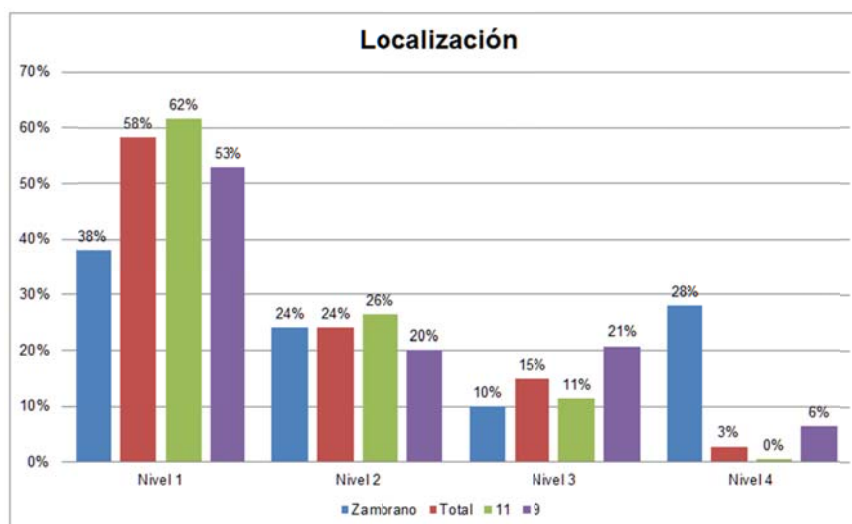


Gráfico 8. Localización

La mayoría de los estudiantes de las poblaciones de estudio involucradas en esta comparación se ubican en el nivel idiosincrásico, se destaca en el nivel uniestructural la población del grado undécimo, en el nivel multiestructural el grado noveno, mientras que en el nivel relacional se acentúa la población del proyecto de Palacios y Zambrano (2009) con una notable diferencia de un 25% de la población total del grupo del presente proyecto.

### Con respecto a la Dispersión

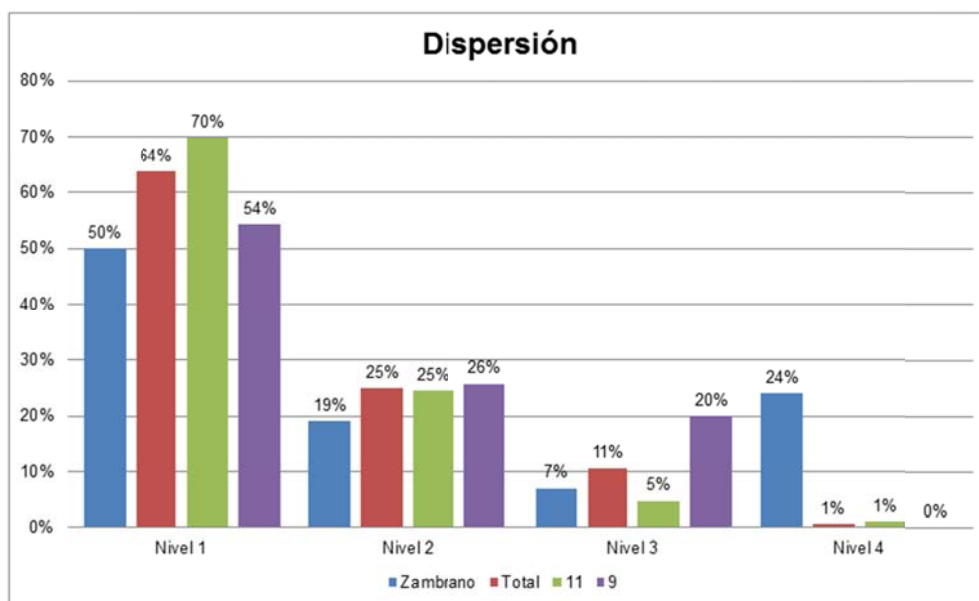


Gráfico 9. Dispersión

El nivel predominante es el idiosincrásico todas las poblaciones comparadas se encuentran entre el 60% y 70%, siendo el grupo de grado undécimo el que registra mayor porcentaje en este nivel. En el uniestructural se observa que los grados noveno y undécimo presentan igual comportamiento. En el nivel multiestructural el porcentaje de estudiantes de grado noveno es cuatro veces el porcentaje del grado undécimo. Los niveles predominantes en la población de estudio del proyecto Palacios y Zambrano (2009) son el idiosincrásico con un 50% y el relacional con un 24%.



### Con respecto a los Intervalos

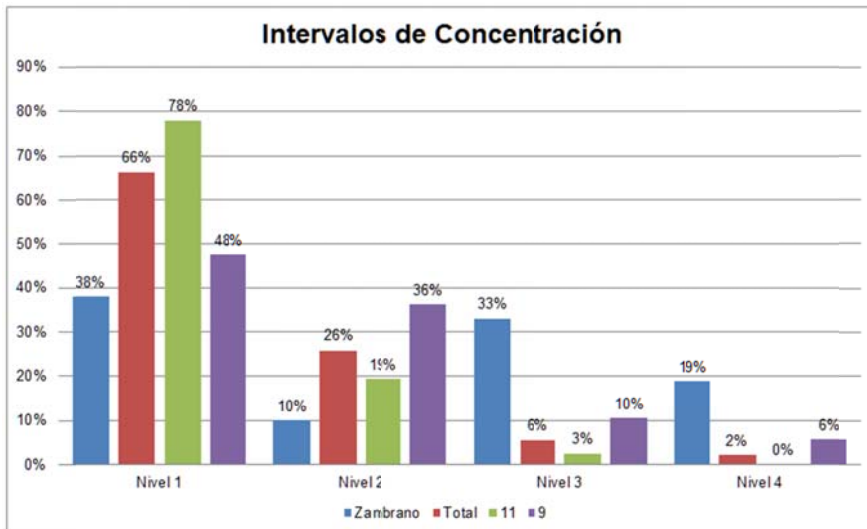


Gráfico 10. Intervalos de Concentración

Se evidencia que el grado noveno presenta un mejor nivel con respecto al grado undécimo. Sin embargo el nivel predominante es el idiosincrásico para todas las poblaciones. En la población de estudio del proyecto Palacios y Zambrano (2009) se destacan los niveles idiosincrásico con un 38% y el multiestructural con un 33%.

### Con respecto a los datos típicos extremos

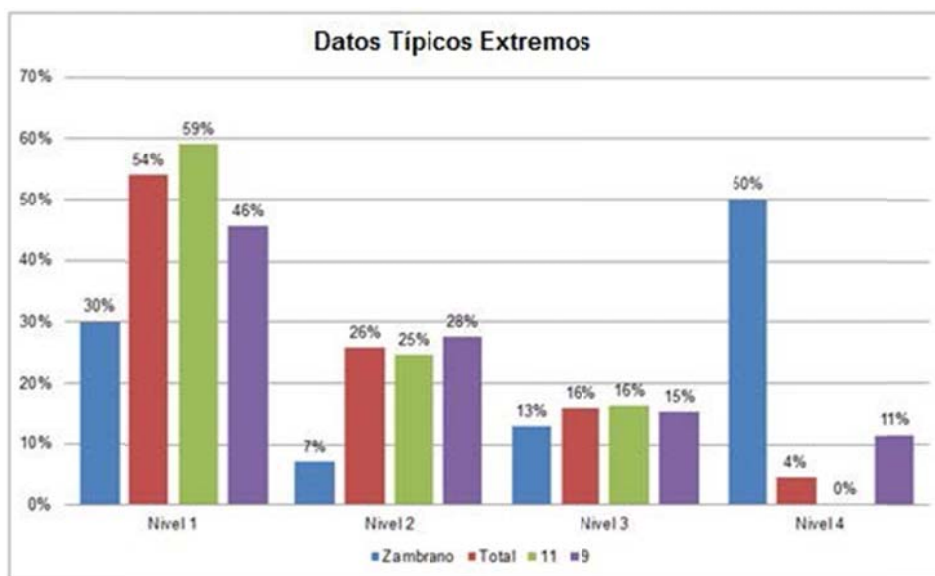


Gráfico 11. Datos Típicos Extremos

Alrededor del 55% de los estudiantes de la población de este estudio se ubican en el nivel idiosincrásico, sin embargo el grupo que registro un puntaje mayor fue el grado undécimo. En los niveles dos y tres los grupos presentan porcentajes similares. Se destaca la población de Palacios y Zambrano (2009) con un 50% en el nivel relacional.

### Con respecto a los Datos Atípicos

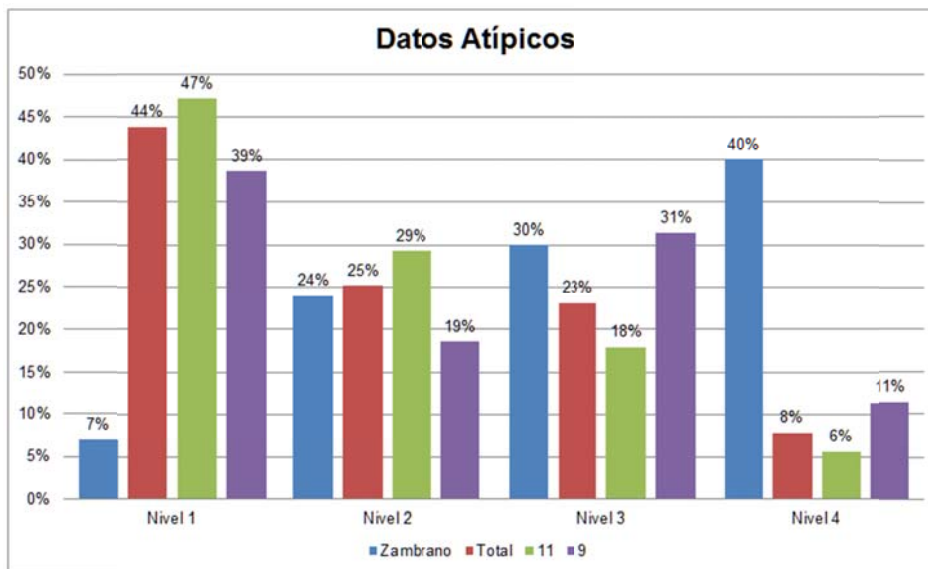


Gráfico 12. Datos Atípicos

Las poblaciones se concentran principalmente entre los niveles 2 y 3 con un porcentaje alrededor del 50%. En los niveles tres y cuatro el grado noveno registra el puntaje más alto con respecto al grado undécimo. Un 70% de la población de Palacios y Zambrano (2009) Se ubican en los dos niveles más altos.

## Con respecto al Total

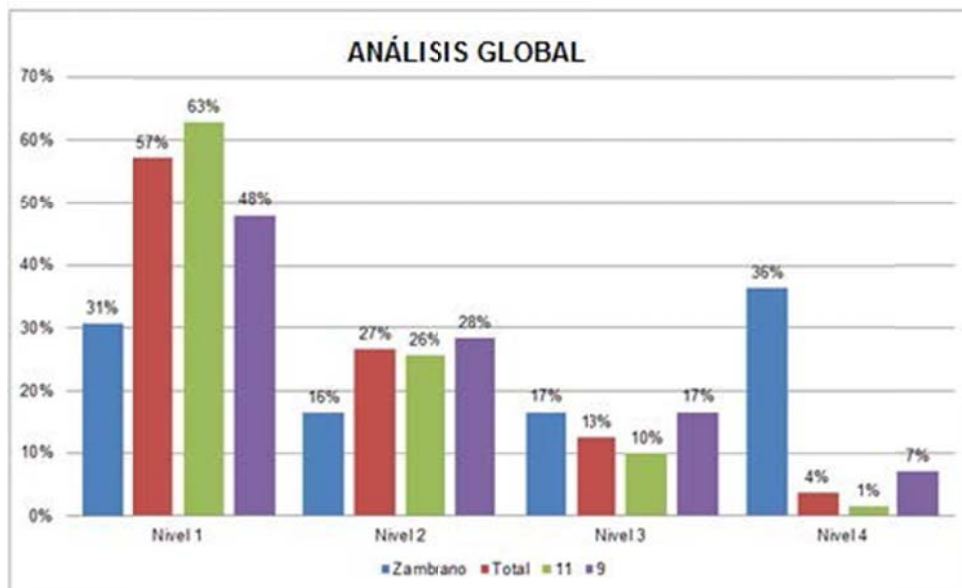


Gráfico 13. Análisis global

Para la población de Palacios y Zambrano (2009) se presenta por orden de frecuencia en primer lugar el nivel IV, luego el nivel I, seguido del nivel III y por último el nivel II.

En los dos grupos analizados de la población de estudio de este proyecto se presenta por orden de frecuencia en primer lugar el nivel I, luego el nivel II, seguido del nivel III y por último el nivel IV. Presentándose una mayor concentración hacia el nivel I por parte de los estudiantes de grado undécimo en comparación a los estudiantes de grado noveno.

El nivel predominante en la población de estudio de este proyecto es el idiosincrásico presentándose con mayor frecuencia en los estudiantes de grado Undécimo. Con respecto a la población del presente proyecto se observa que el grado noveno presenta un razonamiento superior a los estudiantes del grado undécimo. Un 53% de la población de estudiantes de Palacios y Zambrano (2009) se ubican en los niveles Multiestructural y relacional.

A partir de la información obtenida por Palacios y Zambrano (2009) se observa que la población analizada en general posee una apropiación del concepto del gráfico de cajas en comparación a los grupos de estudiantes de este proyecto que se concentraron ampliamente en el nivel I.

## **5. CONCLUSIONES**

### **RESPECTO A LOS OBJETIVOS DEL ESTUDIO**

En este trabajo se logró plantear y orientar a los estudiantes de la población objetivo, con base en una secuencia de actividades que le permitieron acercarse a la identificación de elementos que conforman un gráfico de caja, construirlo y realizar su lectura e interpretación.

Se tomó como punto de partida las categorías propuestas por Palacios y Zambrano (2009), se revisaron y ajustaron al contexto de este trabajo y con base en dicho ajuste se logró identificar y describir aquellos elementos de análisis presentes en un gráfico de caja para caracterizar el razonamiento estadístico de los estudiantes. Por otra parte, el empleo de las categorías utilizadas, posibilitó un proceso de evaluación más preciso y formativo evidenciando de manera metódica y organizada las fortalezas y debilidades del proceso de enseñanza. En particular, se logró la caracterización y análisis del desempeño de los estudiantes al comparar dos distribuciones de datos organizados en gráficos de caja a partir de la cual se deduce la necesidad de revisar las tareas relacionadas con la comparación simultánea de diferentes conjuntos de datos.

### **RESPECTO A LA PROPUESTA DE INSTRUCCIÓN**

Se considera que se debe llevar a cabo un proceso de revisión en lo relacionado a la interpretación simultánea de varios conjuntos de datos representados en gráficos de caja, ya que allí es donde se observa una gran dificultad por parte de los estudiantes al momento de interpretarlos y plantear conclusiones. No obstante, las guías que conforman la secuencia de actividades propuestas para la enseñanza de los gráficos de caja tuvieron una notable aceptación por parte de los estudiantes, dado que se observó el compromiso e interés por parte de ellos para resolverlas.

Por otro lado, el tiempo planeado para el desarrollo de estas guías no fue suficiente, haciéndose necesario abordar algunas de ellas en un tiempo mayor al previsto; una de las razones, pueden ser porque la población de estudio no había tenido una formación

estadística secuencial que le facilitara relacionar los preconceptos con la nueva información.

### **RESPECTO A LOS NIVELES DE CATEGORIZACIÓN DEL RAZONAMIENTO**

La evaluación del razonamiento y pensamiento estadístico se hizo con base en el cuestionario de evaluación diseñado por Palacios y Zambrano (2009). Las respuestas de los estudiantes a este cuestionario fueron el punto de partida para iniciar la tarea de caracterizar el razonamiento y pensamiento estadístico al abordar la clasificación de las respuestas obtenidas, de acuerdo a cada uno de los elementos de razonamiento propuestos en la categorización del modelo SOLO, en los niveles: Preestructural, Uniestructural, Multiestructural y Relacional. Sin embargo para llevar a cabo este proceso fue necesario realizar un ajuste y validación a algunas categorías, ya que varias de las respuestas de los estudiantes no se lograban ubicar con claridad en una categoría específica. Se considera que el ajuste realizado a dichas categorías, es un aporte de este trabajo que mejora y hace más clara la interpretación de respuestas de los estudiantes, en el sentido de facilitar la identificación de las fortalezas, dificultades y falencias que se presentan en el proceso de enseñanza – aprendizaje de los gráficos de caja.

### **RESPECTO AL INSTRUMENTO DE EVALUACIÓN**

El cuestionario de evaluación propuesto por Palacios y Zambrano (2009) requiere de algunos ajustes que le permitan al estudiante observar con claridad la información presentada, como es el caso de los gráficos de caja que representan las dos situaciones con las que se caracteriza el razonamiento estadístico de los estudiantes. Por ejemplo, en las gráficas que representan el peso de hombres y mujeres se evidenció que la ubicación de los cuartiles no es clara, generándose muchas dudas en el valor que corresponde a la mediana del gráfico de caja de la población de mujeres; en este sentido, no era claro si extremo inferior del bigote del conjunto de hombres era igual al de la mediana de las mujeres; la variedad de respuestas de los estudiantes, al respecto fue una evidencia de este problema. De igual forma sucedió con la situación de los usuarios de telefonía celular, ya que en los bigotes inferiores no es claro si inician desde cero u otro valor. Por tanto se recomienda mejorar este aspecto haciendo uso de una recta numérica que indique los valores más claramente.

Se propone adicionar a los gráficos planteados, una línea que indique los límites de la zonas de anomalías, al igual plantear interrogantes alrededor de está de tal manera que genere dudas en el estudiante y le permita reflexionar más acerca de los datos típicos y atípicos.

Para ser aplicado el instrumento de evaluación a una población mayor, o de manera masiva, es necesario reformular las preguntas del instrumento con opciones cerradas de respuestas que faciliten su tabulación y consolidación; si bien es cierto las preguntas abiertas dan mayor información que las preguntas de opción cerrada; su valoración se hace difícil al momento de asignar el nivel correspondiente quedando sujeto al criterio del evaluador.

### **RESPECTO A LOS RESULTADOS**

En el presente estudio se observa que los estudiantes de grado noveno de la IED Técnico Comercial Mariano Ospina Rodríguez, presentan un mejor nivel de razonamiento estadístico en la mayoría de los elementos de análisis establecidos, comparado con los estudiantes de grado undécimo de la IED General Carlos Albán, que fueron los dos grupos de estudio considerados. Esta situación parece coherente con el hecho de observar un mayor nivel académico de la IED Técnico Comercial Mariano Ospina Rodríguez, registrado según el ICFES en un nivel alto, en comparación con el registrado por la IED General Carlos Albán, registrado en un nivel medio. Por lo tanto no se confirmó el supuesto implícito, de que a mayor nivel escolar, mayor nivel de razonamiento estadístico.

De acuerdo al análisis que se llevó a cabo, los elementos de razonamiento estadístico que reflejan un nivel idiosincrásico de los estudiantes analizados en el presente proyecto en su orden son: Tamaño de muestra, localización, intervalos de concentración, escala, dispersión, datos atípicos y datos típicos extremos. Mientras que para la Población de estudio del proyecto Palacios y Zambrano (2009) son en su orden: Dispersión, Localización, Intervalos de Concentración, Tamaño de la muestra, Datos adyacentes (datos típicos extremos), Escala y Datos Atípicos.

El elemento de análisis que presentó mayor dificultad para la población de estudio de Palacios y Zambrano (2009) fue la dispersión, mientras que para la población analizada

en el presente trabajo, se evidencia que este elemento no revistió igual nivel de dificultad siendo este uno de los de mayor grado de comprensión por parte de los estudiantes.

Por otra parte, el tamaño de la muestra es el elemento en el cual la población de interés presentó mayor dificultad dado que lo relacionaron con la amplitud del gráfico, mostrándose una comprensión errada de este concepto, por tanto es importante en el momento de orientar su interpretación hacer hincapié acerca de los errores comunes que se pueden presentar en el mismo.

El bajo nivel de razonamiento estadístico de los estudiantes puede ser atribuido a que no han recibido educación estadística constante, siendo para algunos de ellos la aplicación de las guías propuestas un primer acercamiento al conocimiento de la estadística.

En la ejecución del proyecto no se aplicó ningún tipo de muestreo intencionado, siendo esta una de las posibles razones por las cuales los resultados obtenidos en la población de estudio del proyecto de Palacios y Zambrano arrojaron un nivel de razonamiento estadístico mucho más alto que el de la población de estudio del presente proyecto. De hecho los cuestionarios que no registraron respuestas coherentes o que sencillamente tuvieron preguntas no respondidas entraron en el análisis siendo estas categorizadas en el nivel idiosincrásico.

## REFERENCIAS

- Batanero, C; Estepa, A; Díaz, G. (1991). Análisis Exploratorio de Datos: sus posibilidades en la enseñanza secundaria. *Suma*, nº 9, 1991: 25-31.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64-83, en <http://www.stat.auckland.ac.nz/serj>
- Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistics learning. *Mathematical Thinking and Learning*, 2 (1 y 2), 127 – 155
- Ben-Zvi, D y Garfield, J (2004) The Challenge of Developing Statistical Literacy, Reasoning and Thinking. Kluwer Academic Publishers. Printed in the Netherlands. P.p 47–78.
- Biggs, J y Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Biggs, J y Collis, K. (1991). Multimodal Learning and the Quality of Intelligent Behavior, en Rowe, H. (ed.) *Intelligence Reconceptualization and Measurement*. LEA, Australian Council for Educational Research, pp. 57-76.
- Cobo, B. (2003). Significado de las medidas de posición central para los estudiantes de Secundaria. Universidad de Granada. Departamento de Didáctica de la Matemática.
- Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5), 382-393.
- Freixa, I. (1992). Análisis exploratorio de datos: Nuevas Técnicas Estadísticas. Promociones y publicaciones Universitarias S.A. Barcelona.
- Garfield, J. (2002) The Challenge of Developing Statistical Reasoning" *Journal of Statistics Education* [Online], 10(3). En [www.amstat.org/publications/jse/v10n3/garfield.htm](http://www.amstat.org/publications/jse/v10n3/garfield.htm).
- Gómez, P. (2002). Análisis didáctico y diseño curricular en matemáticas. *Revista EMA*, 7(3), 251-292.



Jones, G. Langrall, C. Money, E. y Thornton, C. (2004). Models of development. En D. Ben-Zvi y J. Garfield (eds), *The challenge of thinking*, pp 97 -117. Dordrecht: Kluwer Academic Publishers.

Medina, A. (1998). *Didáctica de las matemáticas: Estrategias Metodológicas para la Enseñanza y el Aprendizaje de las Matemáticas*. Duitama: Universidad Pedagógica y Tecnológica de Colombia.

Ministerio de Educación Nacional (1998). *Lineamientos curriculares de matemáticas*. Bogotá: [MEN].

Ministerio de Educación Nacional (2006). *Estándares Básicos de Competencias en Matemáticas*. Bogotá: Ministerio de Educación Nacional [MEN].

Peeg, J. Gutiérrez, A. y Huerta, M. (1997). Assessing Reasoning Abilities in Geometry, en Villani, V. y Mammana, C. (eds.) (1998), *Perspectives on the teaching of Geometry for the 21st Century*. Colección: Publicaciones del ICMI, Kluwer Academic Press, pp. 275-295.

Pfannkuch, M. (2006). Comparing box plot distributions: teacher's reasoning. The University of Auckland, New Zealand [m.pfannkuch@auckland.ac.nz](mailto:m.pfannkuch@auckland.ac.nz) *Statistics Education Research Journal*, 5(2), 27-45, en: <http://www.stat.auckland.ac.nz/serj>. *International Association for Statistical Education (IASE/ISI)*.

Tukey, J. (1977) *Exploratory Data Analysis*. Editorial Addison Wesley.

Velleman, P. y Hoaglin D. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press.

Wild, C. J. y Pfannkuch, M. (1999), Statistical Thinking in Empirical Enquiry, *International Statistical Review*, 67, pp. 223 –265. Traducido al castellano por el Dr. Román Hernández Martínez.

Zambrano, W. y Palacios, R. (2009). *Razonamiento estadístico de estudiantes de secundaria al comparar conjunto de datos representados mediante gráficos de caja* (Tesis de maestría). Bogotá: Universidad Pedagógica Nacional.

# ANEXOS

## ANEXO 1

### MEDIDAS DE TENDENCIA CENTRAL

|  |                   |                                    |  |
|--|-------------------|------------------------------------|--|
| IED TÉCNICO COMERCIAL MARIANO OSPINA RODRÍGUEZ<br>IED GENERAL CARLOS ALBAN   |                   |                                    |  |
| GUIA: 1  | ÁREA: Matemáticas | TEMA: Medidas de tendencia central |  |
| NOMBRE:  | FECHA:            | GRADO: _                           |  |
| DESEMPEÑO: Reconoce el orden en la recta numérica e identifica las medidas de tendencia central de un conjunto de datos.   |                   |                                    |  |
| REVISANDO CONCEPTOS PREVIOS: Defina con sus propias palabras los siguientes conceptos:<br>Recta numérica: _____<br>Orden en los números reales: Moda: _____<br>Media aritmética: Mediana _____   |                   |                                    |  |
| SITUACIÓN PROBLEMA: <u>La llegada de los estudiantes a la Institución</u><br>Los siguientes datos corresponden al tiempo (en minutos) de llegada de 9 estudiantes a la Institución Educativa. Suponga que las 7:00 a.m. corresponde al punto cero en la recta numérica y los valores negativos, al tiempo de los estudiantes que llegan antes de las 7:00 a.m. y valores positivos al tiempo de los estudiantes que llegan después de las 7:00 a.m.<br>Ubica los datos en la recta numérica. |                   |                                    |  |
| <table border="1"><tr><td>-5, -15, 12, -25, -18, 4.5, 3, -2, -15</td></tr></table>   |                   |                                    | -5, -15, 12, -25, -18, 4.5, 3, -2, -15 |
| -5, -15, 12, -25, -18, 4.5, 3, -2, -15   |                   |                                    |  |
| FORMULACIÓN  |                   |                                    |  |
| Identifica el dato que más se repite. _____  |                   |                                    |  |
| ¿Cómo se llama este dato? _____  |                   |                                    |  |
| ¿Qué representa este dato en la situación anterior? _____  |                   |                                    |  |
| c. Halla el promedio del conjunto de los datos e interpreta la respuesta: _____  |                   |                                    |  |
| d. Organiza los datos de menor a mayor. _____  |                   |                                    |  |
| e. Toma el dato del centro. ¿Qué considera que significa este valor? _____   |                   |                                    |  |
| f. ¿Qué relación tendrán los resultados hallados en los encisos b, c y e? _____  |                   |                                    |  |

g. ¿Qué indican los resultados negativos acerca de la situación?\_

h. Hay un estudiante que llegó a las 7:02 minutos y no fue tenido en cuenta. Incluya el nuevo dato en los datos ordenados, proponga una estrategia para encontrar el valor del dato central y argumente que significa este dato.

SOCIALIZACIÓN: En grupos de cuatro estudiantes se confrontarán los resultados obtenidos. Cada integrante justifica su respuesta y el grupo llegará a un consenso para sustentarlo en plenaria. El docente realiza el correspondiente acompañamiento para aclarar las dudas y llegar a una formalización.

### FORMALIZACIÓN

El valor que más se repite dentro del conjunto de datos se denomina: MODA y se simboliza ( $M_o$ ).

MEDIA ARITMÉTICA O PROMEDIO  $\bar{x} = \frac{\text{Suma de todos los valores observados}}{\text{Número de observaciones}}$

Dado un conjunto numérico de datos,  $x_1, x_2, \dots, x_n$ , se define su media aritmética como

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

MEDIANA ( $M_e$ ): La mediana representa el valor de la variable de posición central en un conjunto de datos ordenados. Los datos menores o iguales que la mediana representan el 50% de los datos y los que sean mayores que la mediana representan el otro 50% del total de datos.

Para calcularla

Primero se ordenan los datos de menor a mayor o viceversa

Si el número de datos es par, se toman los dos valores centrales y se obtiene sacando el promedio entre ellos.

$$M_e = \frac{(x_{n/2} + (x_{n/2+1}))}{2}$$

Si el número de datos es impar el valor central es la mediana.  $M_e = x_{(n+1)/2}$

### APLICA LO APRENDIDO

1. Cinco amigos del grado once salen a comer helado, tres compraron helado de \$3.500 y en promedio cada uno canceló \$4.100. Completa la siguiente tabla y encuentra las medidas de tendencia central.

2. Un entrenador de pista debe decidir cuál de los dos corredores debe elegir para la próxima carrera de 100 metros planos. El entrenador basará su decisión en los resultados que tiene registrados de cinco carreras entre los dos atletas, realizadas en un período de una hora con descanso de 15 segundos.

| Amigos  | Valor de cada helado |
|---------|----------------------|
| Juan    |                      |
| Ricardo |                      |
| Tania   |                      |
| Luz     |                      |
| Miguel  |                      |

| Atleta | Tiempo /número de carreras |      |      |      |      |
|--------|----------------------------|------|------|------|------|
|        | 1                          | 2    | 3    | 4    | 5    |
| Juan   | 12,1                       | 12   | 12   | 16,8 | 12,1 |
| Omar   | 12,3                       | 12,4 | 12,4 | 12,5 | 12,4 |

a. Sobre la base de los datos anteriores, halle la media aritmética. ¿Cuál de los dos corredores debe elegir y por qué?

b. ¿Debería ser diferente la selección si el entrenador supiera que el atleta Juan se cayó al principio de la cuarta carrera? Fundamente con la

mediana.

- c. Comente la diferencia entre los conceptos media aritmética y mediana como medidas de tendencia central y su relación con los incisos a. y b.
- d. Con base en la moda, media aritmética y la mediana ¿cuál atleta recomendaría al entrenador para que fuera seleccionado?
3. Se preguntó a 13 estudiantes por el número de veces que ha asistido al departamento de orientación. Los resultados son: 0, 3, 2, 5, 0, 3, 7, 1, 2, 0, 5, 3, 4. Calcula los valores de tendencia central e interpreta estos resultados.

**ANEXO 2  
CUARTILES**

IED TÉCNICO COMERCIAL MARIANO OSPINA RODRÍGUEZ  
IED GENERAL CARLOS ALBAN  
GUIA: 2 TEMA: CUARTILES

NOMBRE: \_\_\_\_\_ FECHA: \_\_\_\_\_ GRADO: \_\_\_\_\_  
INTRODUCCIÓN:

Para el análisis de los datos además de utilizar las medidas de tendencia central (media, mediana y moda) es útil dividir el conjunto de datos en cuatro partes iguales denominadas cuartiles.

**REVISANDO CONCEPTOS PREVIOS**

1. Relacione cada uno de los siguientes términos con su respectivo concepto:
- 1.1 Media Aritmética (Promedio) ( )      a. Es el dato más repetido, el valor de la variable con mayor frecuencia absoluta.
  - 1.2 Moda ( )      b. Valor de la variable que deja por debajo de sí a la mitad de los datos, una vez que éstos están ordenados.
  - 1.3. Mediana ( )      c. Valor de la variable que se encuentra a igual distancia de todos los datos.

**Situación problema 1:**

Un objeto pequeño se pesó con un mismo instrumento, separadamente por nueve estudiantes en una clase de ciencias. Los pesos obtenidos por cada estudiante (en gramos) se muestran a continuación:

6.2 6.0 6.0 15.3 6.1 6.3 6.2 6.15 6.2

Los estudiantes quieren determinar con la mayor precisión posible el peso real del objeto. ¿Cuál de los siguientes métodos les recomendarías usar y por qué?

- a) Usar el número más común, que es 6.2
- b) Usar 6.15, puesto que es el dato más preciso
- c) Sumar los 9 números y dividir la suma por 9
- d) Desechar el valor 15.3; sumar los otros 8 números y dividir por 8

Determinación de los cuartiles

Situación 2: Análisis de la estatura en cm de un grupo de 15 estudiantes

El cálculo de los cuartiles, que implica dividir el conjunto de datos en cuatro grupos de datos, se puede realizar siguiendo estos pasos: Se tiene la información acerca de la estatura en cm de 15 estudiantes:

|            |      |      |      |       |        |        |        |        |       |          |       |       |       |       |        |
|------------|------|------|------|-------|--------|--------|--------|--------|-------|----------|-------|-------|-------|-------|--------|
| Estudiante | Juan | José | John | Jaime | Andrés | Carlos | Miguel | Felipe | David | Santiago | Oscar | Fredy | Diego | Mauro | Steven |
| Estatura   | 145  | 162  | 152  | 168   | 169    | 151    | 153    | 153    | 147   | 150      | 142   | 151   | 156   | 163   | 150    |

**Paso 1 Ordene la estatura de los estudiantes de menor a mayor**

|          |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| Posición | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Estatura |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |

A partir de la información de los estudiantes ordenada de menor a mayor:

**Paso 2 Halle la mediana de estos datos ¿Cómo se interpreta este resultado?**

**Paso 3 Halle la mediana de los datos 1 al 7 ¿Cómo se interpreta este resultado?**

**Paso 4 Halle la mediana de los datos 9 al 15 ¿Cómo se interpreta este resultado?**

Situación 3: Análisis de la estatura de un grupo de 16 estudiantes

Suponga que ingresa un nuevo estudiante al grupo el cual tiene una estatura de 150 cm.

**Paso 1 Ordene la estatura de los mismos de menor a mayor**

|          |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|
| Posición | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Estatura |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |

A partir de la información de los estudiantes ordenada de menor a mayor:

Paso 2 Halle la mediana de estos datos ¿Cómo se interpreta este resultado?\_

Paso 3 Halle la mediana de los datos 1 al 8 ¿Cómo se interpreta este resultado?\_

Paso 4 Halle la mediana de los datos 9 al 16 ¿Cómo se interpreta este resultado?\_

#### FORMALIZACIÓN

Si se tiene un conjunto de valores  $X_1, X_2, X_3 \dots X_n$ , que en nuestro ejemplo son las estaturas de los estudiantes ordenadas de menor a mayor se requiere saber los valores que dividen el conjunto de datos en cuatro partes iguales, a estos valores se les denominan cuartiles.

Los cuartiles dividen la distribución de los datos en cuatro partes iguales y cada una de estas partes es igual al 25% del total de los datos.

El primer cuartil: El primer cuartil Q1, es aquel valor de la variable que supera la cuarta parte o el 25% de las observaciones y es superado por las tres cuartas partes o el 75% de las observaciones.

Este dato es el que estábamos localizando en el paso 3 de las situaciones propuestas

El Segundo cuartil: (coincide, es idéntico o similar a la mediana,  $Q2 = Me$ ), es aquel valor de la variable que supera la mitad o el 50% de las observaciones y es superado por la mitad o el 50% de las observaciones.

Este dato es el que estábamos localizando en el paso 2 de las situaciones propuestas

El tercer cuartil: El tercer cuartil Q3 es aquel valor de la variable que supera las tres cuartas partes o el 75% de las observaciones y es superado por la cuarta parte o el 25% de las observaciones.

Este dato es el que estábamos localizando en el paso 4 de las situaciones propuestas

#### APLICA LO APRENDIDO

1. En una prueba de atletismo se registran los tiempos obtenidos por 20 participantes

|              |           |          |       |          |       |          |        |         |         |         |
|--------------|-----------|----------|-------|----------|-------|----------|--------|---------|---------|---------|
| Participante | Tatiana   | Angélica | Tania | Lina     | Lisa  | Carolina | Lorena | Paula   | Milena  | Daniela |
| Tiempo (min) | 10.03     | 14.16    | 14.85 | 9.3      | 5.23  | 19.14    | 14.25  | 5.26    | 17.34   | 13.8    |
| Participante | Valentina | Isabella | Angie | Jennifer | Luna  | Mary     | Paola  | Julieth | Vanessa | Nancy   |
| Tiempo (min) | 9.91      | 19.32    | 5.01  | 14.99    | 12.18 | 8.57     | 18.37  | 19.6    | 11.66   | 18.44   |

- Calcule los cuartiles Q1, Q2 y Q3 correspondientes a los tiempos registrados en la prueba de atletismo
- ¿Cuál es el tiempo que no puede superar la cuarta parte de los participantes en la prueba?
- ¿Cuál es el tiempo que no puede superar la mitad parte de los participantes en la prueba?
- ¿Cuál es el tiempo que no puede superar las tres cuartas partes de los participantes en la prueba?

2. A continuación se registran las notas obtenidas por 13 estudiantes en las asignaturas de Matemáticas y Ciencias

|             |      |      |      |       |        |        |        |        |       |          |       |        |       |
|-------------|------|------|------|-------|--------|--------|--------|--------|-------|----------|-------|--------|-------|
| Estudiante  | Juan | José | John | Jaime | Andrés | Carlos | Miguel | Felipe | David | Santiago | Oscar | Freddy | Diego |
| Matemáticas | 24   | 29   | 27   | 21    | 25     | 20     | 11     | 22     | 38    | 24       | 15    | 31     | 14    |
| Ciencias    | 28   | 47   | 21   | 45    | 36     | 28     | 38     | 23     | 21    | 30       | 49    | 40     | 31    |

- Calcule los cuartiles Q1, Q2 y Q3 correspondientes a las notas de matemáticas
- Calcule los cuartiles Q1, Q2 y Q3 correspondientes a las notas de ciencias
- ¿Cuál es la nota de matemáticas que no puede superar el 25% de los estudiantes?
- ¿Cuál es la nota de matemáticas que no puede superar el 50% de los estudiantes?
- ¿Cuál es la nota de matemáticas que no puede superar el 75% de los estudiantes?
- ¿Cuál es la nota de ciencias que puede superar el 25% de los estudiantes?
- ¿Cuál es la nota de ciencias que puede superar el 50% de los estudiantes?
- ¿Cuál es la nota de ciencias que puede superar el 75% de los estudiantes?

ANEXO 3  
CUARTILES

IED TÉCNICO COMERCIAL MARIANO OSPINA RODRÍGUEZ  
IED GENERAL CARLOS ALBAN

GUIA: 3 TEMA: Construcción del gráfico de caja ÁREA: Matemáticas  
NOMBRE: \_ FECHA: \_ GRADO: \_

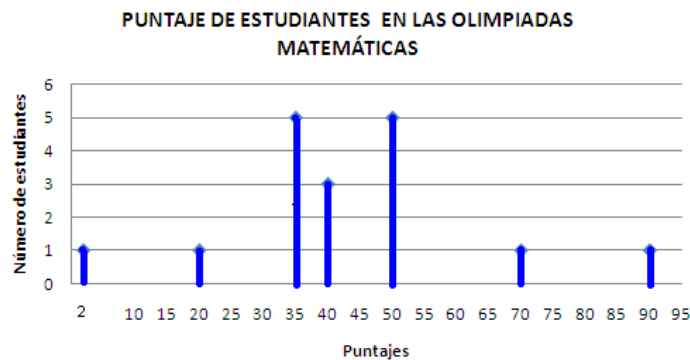
INTRODUCCIÓN:

Para representar datos estadísticos se utilizan diversos gráficos que facilitan la observación e interpretación de la información, entre ellos encontramos el denominado “diagrama de caja y bigotes”.

- ◆ ¿Qué tipos de gráficos de estadística conoce?\_
- ◆ Dé un ejemplo de los gráficos estadísticos que conoce:

Situación problema:

El siguiente gráfico representa el puntaje obtenido por 17 estudiantes en las olimpiadas matemáticas (valoradas con un puntaje de 1 a 100).



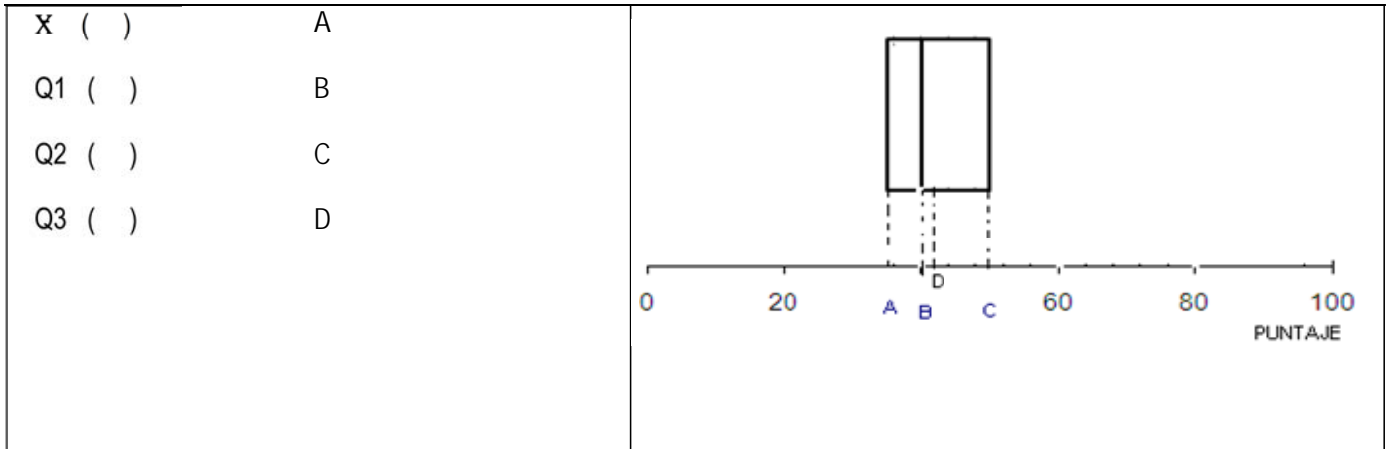
1. Ubica en la siguiente tabla y de manera ordenada los 17 puntajes obtenidos por los estudiantes.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|

2. Determina las medidas estadísticas de promedio y cuartiles y ubícalas en el diagrama de barras

X= \_ Q1 = \_ Q2 = \_ Q3 = \_

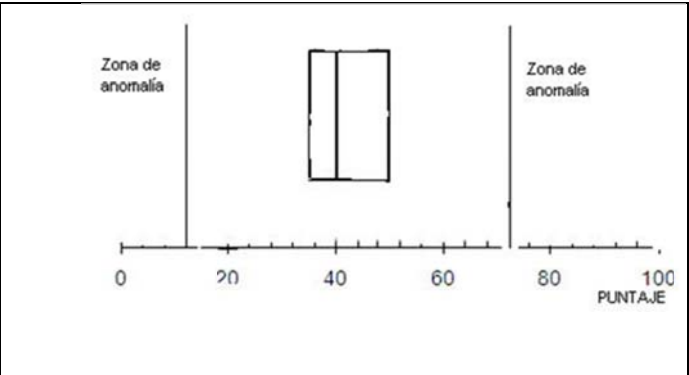
3. Elabore el siguiente gráfico (en una hoja) y relacione las medidas estadísticas halladas anteriormente y los valores que indican las letras A, B, C y D.



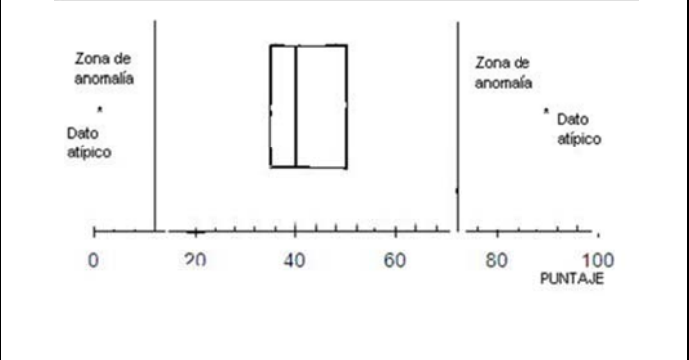
Posteriormente calcula los siguientes datos

- Diferencia  $(Q3 - Q1)$  Denominado rango intercuartilico
- A Q1 réstale  $1.5(Q3 - Q1)$  \_
- A Q3 súmale  $1.5(Q3 - Q1)$  \_

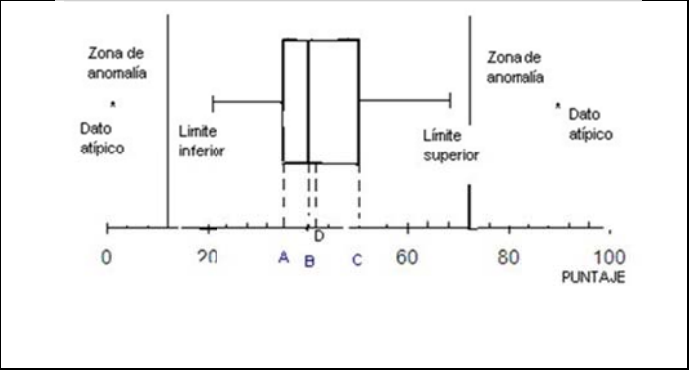
Sobre el gráfico que se está elaborando, se traza a cada lado una recta vertical que pase por los puntos obtenidos en los incisos b y c. De esta forma quedan definidas las zonas de anomalías.



4. Observa si quedaron algunos datos por fuera de este intervalo, ubíquelos guiándose por la recta numérica marcándolos con un asterisco, como se observa en la figura. Estos datos se denominan datos atípicos.



5. Ubica el dato menor no atípico y el dato mayor no atípico y traza una línea paralela a la recta numérica desde el rectángulo hasta dichos datos como se observa en la figura. Estos datos se denominan límites inferior y superior.



◆ Milena dice que los puntajes 10 y 70 puntos son valores atípicos ¿usted está de acuerdo con ella? Si \_ No \_  
¿Porqué?\_

◆ Según la información del gráfico se podría afirmar que la mayoría (más del 50%) de los puntajes supera el puntaje de B?



Si \_ No\_ ¿Porqué?\_

- Si Juan obtiene un puntaje mayor que Q3, se puede afirmar que:
  - Se encuentra en el 25% de los puntajes más bajos.
  - Juan obtuvo un puntaje igual al promedio
  - Se encuentra en el 25% de los puntajes más altos.
  - Juan es un estudiante de bajo rendimiento.

¿Se puede afirmar que el puntaje promedio obtenido por los estudiantes en las olimpiadas matemáticas es menor que Q2?

Si NO\_ ¿Porqué?\_

### FORMALIZACIÓN

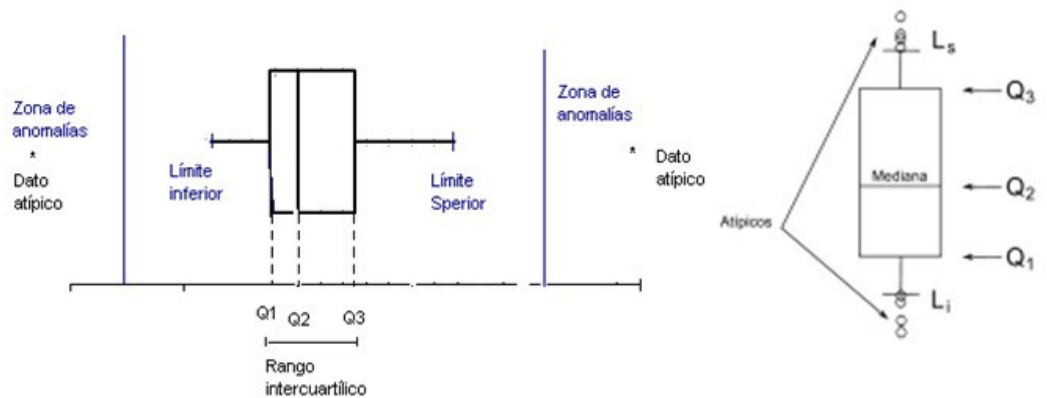
El proceso de construcción del diagrama de caja se lleva a cabo de la siguiente manera:

- La caja se construye con base en los valores de los cuartiles, ubicándolos en la recta numérica y se dibuja un rectángulo cuya base es la diferencia entre Q3 y Q1 (Rango intercuartílico).
- Se dibuja una línea para dividir la caja con base en el valor que representa Q2 (La mediana).
- Se trazan dos rectas verticales que pasan por los puntos  $[Q1 - 1.5(Q3 - Q1)]$  y  $[Q3 + 1.5(Q3 - Q1)]$  que son las que definen las zonas de anomalías.
- Se traza una línea paralela a la recta numérica desde el Q1 hasta el dato menor que no pertenezca a la zona de anomalías, al igual que una línea paralela desde Q3 hasta el dato mayor que no pertenezca a la zona de anomalías. Estos datos corresponden a los límites inferior y superior.
- Los datos que se encuentran en la zona de anomalías se llaman datos atípicos, se marcan con un asterisco y son valores que influyen significativamente en el cálculo de las medidas que caracterizan la variable.
- Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

El diagrama de caja es un resumen gráfico en el que se describen varias de las características más destacadas de un conjunto de datos. En el proceso de interpretación se pueden determinar:

- Valores atípicos
- Valores de los cuartiles
- Ubicación del promedio

En los siguientes gráficos se observa un diagrama de caja de forma horizontal y uno de forma vertical.



### APLICA LO APRENDIDO

1. Situación problema: Los siguientes datos corresponden a los puntajes obtenidos en las pruebas de admisión por 25 aspirantes a ingresar a la Universidad Nacional.

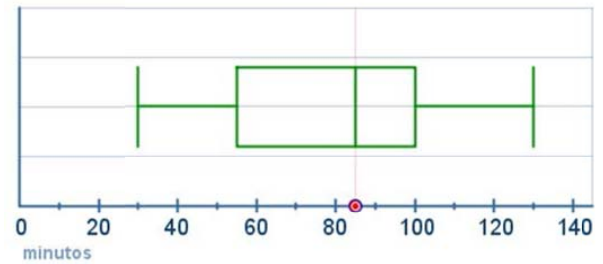
55, 44, 64, 76, 20, 41, 56, 75, 46, 50, 41, 49, 56, 51, 54, 62, 54, 55, 54, 76, 59, 57, 42, 47, 92

- Elabora el diagrama de caja
- Identifica si hay datos atípicos y argumenta que significan estos datos.
- ¿Qué puntaje obtuvo el 75% de los estudiantes?

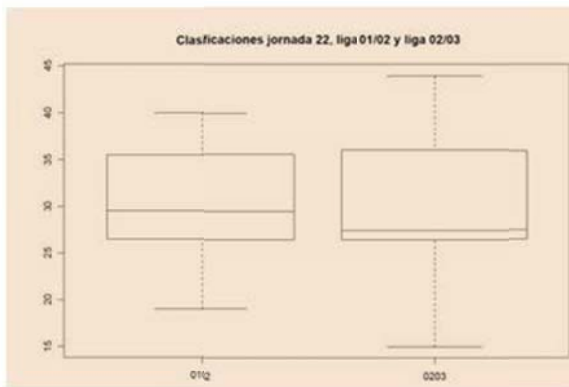
- d. ¿El 50% de los estudiantes que puntaje obtuvieron?
- e. Si se encuentran datos atípicos, ¿considera que la media aritmética es un buen representante del puntaje promedio obtenido por los estudiantes? Si no lo es, ¿Qué propone para que está sea representativa?

2. Situación problema: El siguiente gráfico de caja muestra los minutos que tarda en hacer efecto un medicamento en una población. Interpreta la información que se presenta y responde a las preguntas.

- a. ¿A qué porcentaje de la población había hecho efecto al cabo de 30 minutos?
- b. ¿Al cabo de cuántos minutos había hecho efecto al 50 % de la población?
- c. ¿Cuántos minutos tardó en hacer efecto al 100% de la población?
- d. ¿A qué porcentaje había hecho efecto a los 55 minutos?
- e. ¿Cuánto tardó en hacer efecto a las tres cuartas partes de la población?



3. Situación problema: La puntuación de los equipos de la liga de la temporada 01/02 y 02/03 en primera división se pueden comparar en un diagrama de caja y bigotes.



- a. Encuentra los valores de la mediana, los Q1y Q3, limites superiores e inferiores.
- b. Realiza una conclusión sobre lo que se puede inferir al comparar los diagramas.

ANEXO 4  
ANÁLISIS E INTERPRETACIÓN DE LOS GRÁFICOS DE CAJA.

IED TÉCNICO COMERCIAL MARIANO OSPINA RODRÍGUEZ  
IED GENERAL CARLOS ALBAN  
GUIA: 4 ANÁLISIS E INTERPRETACIÓN DE LOS GRÁFICOS DE CAJA.

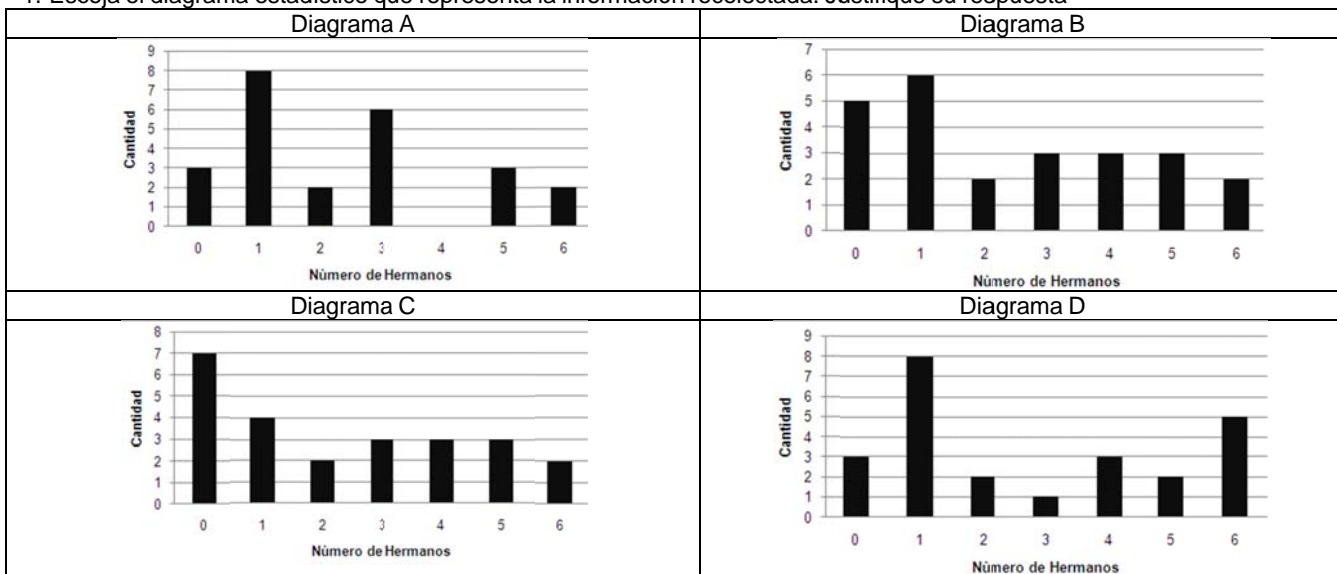
NOMBRE: \_\_\_\_\_ FECHA: \_\_\_\_\_ GRADO: \_\_\_\_\_ TIEMPO: \_\_\_\_\_

**EJEMPLO 1**

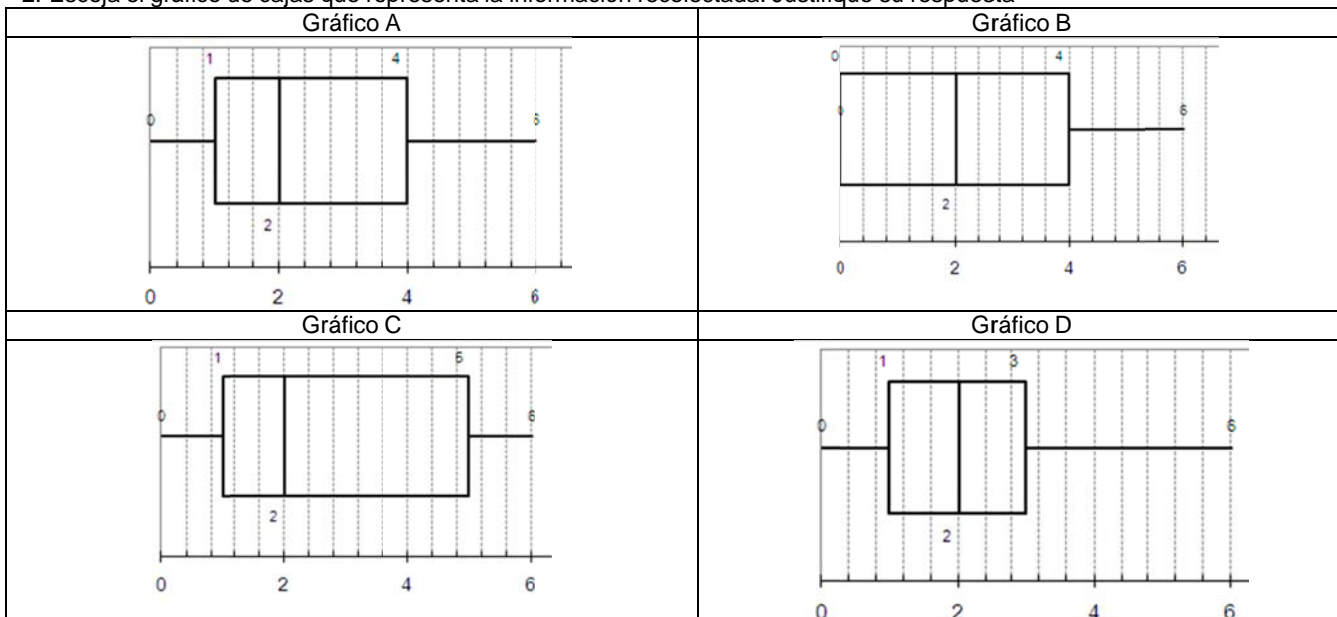
Se tiene la siguiente información acerca del número de hermanos de 24 personas

|        |         |         |         |        |          |       |         |       |           |        |        |
|--------|---------|---------|---------|--------|----------|-------|---------|-------|-----------|--------|--------|
| Miguel | Juanita | Tatiana | Lina    | Julián | Andrés   | Paola | Sandra  | Pablo | José      | Camilo | Pedro  |
| 3      | 1       | 4       | 0       | 2      | 1        | 3     | 4       | 1     | 0         | 5      | 6      |
| Juan   | Luis    | Carlos  | Johanna | Leydi  | Carolina | Paula | Liliana | María | Alejandro | Daniel | Darwin |
| 1      | 0       | 0       | 4       | 3      | 1        | 0     | 2       | 1     | 5         | 6      | 5      |

1. Escoja el diagrama estadístico que representa la información recolectada. Justifique su respuesta



2. Escoja el gráfico de cajas que representa la información recolectada. Justifique su respuesta



3. A partir del Gráfico de Cajas y el Diagrama de Barras escogido. Los datos están más concentrados a la izquierda o a la derecha? ¿Qué se puede afirmar acerca del número de hermanos que poseen las personas encuestadas? \_\_\_\_\_

4. A partir del Diagrama de Caja y Gráfico de Barras escogidos Determine los cuartiles y los valores mínimo y máximo

Q1 \_\_\_\_\_  
 Q2 (Mediana) \_\_\_\_\_  
 Q3 \_\_\_\_\_  
 Mínimo \_\_\_\_\_  
 Máximo \_\_\_\_\_

5. Ordene del 1 al 4 los siguientes intervalos desde el más concentrado (1) hasta el más disperso (4), recuerde que cada uno de estos intervalos contiene el 25% de los datos.

- a. Mínimo al Q1 ( )
- b. Q1 al Mediana ( )
- c. Mediana al Q3 ( )
- d. Q3 al Máximo ( )

6. Relacione los Gráficos de Barras con los Respectivos Gráficos de Caja, Justifique su respuesta

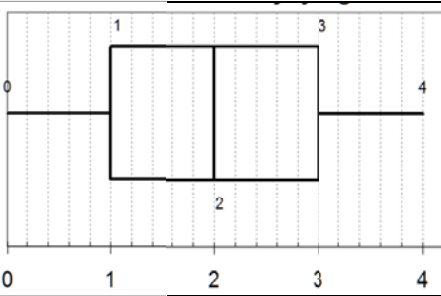
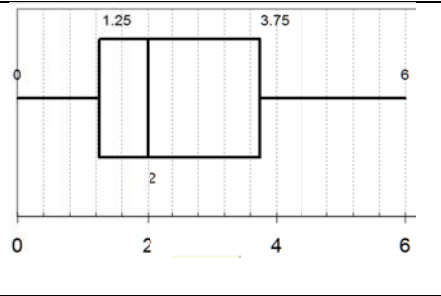
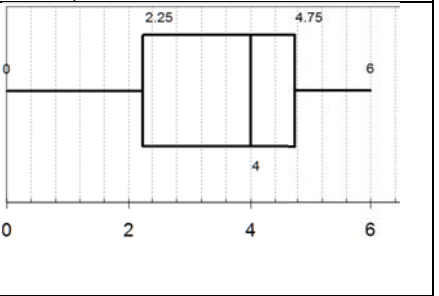
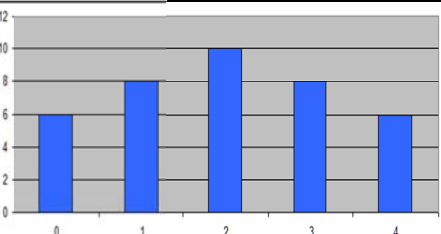
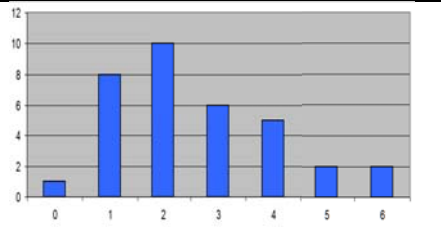
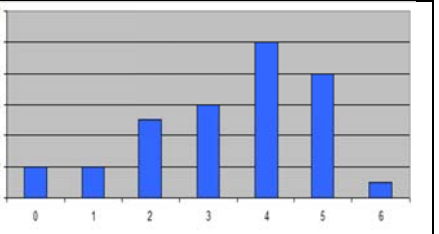
| Diagrama de Barras | Gráficos de Caja |
|--------------------|------------------|
| Diagrama A ( )     | Gráfico A        |
| Diagrama B ( )     | Gráfico B        |
| Diagrama C ( )     | Gráfico C        |
| Diagrama D ( )     | Gráfico D        |

## FORMALIZACIÓN

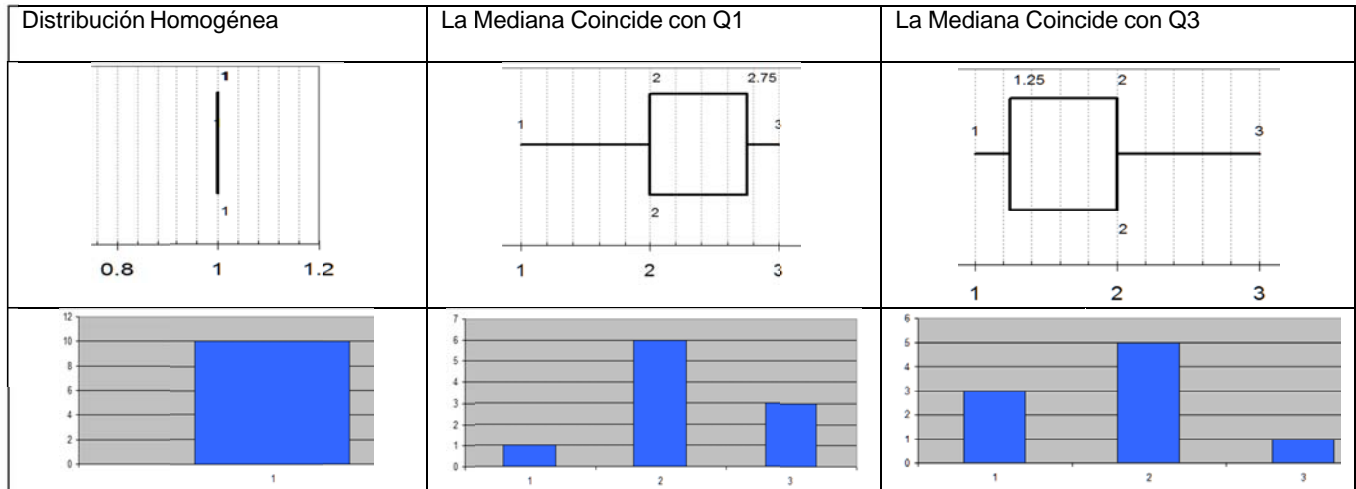
**Tenga en cuenta las siguientes consideraciones a la hora de interpretar el Gráfico de Caja:**

- Mientras más larga sean la caja y los bigotes, más dispersa es la distribución de datos.
- El Gráfico de Cajas no suministra información para establecer el número de datos.
- La línea que representa la mediana (Q2) muestra la simetría del conjunto de datos.
- La interpretación del gráfico está basada en la comparación de las longitudes de los cuatro intervalos. (Mínimo al Q1, Q1 a la Mediana, Mediana al Q3 y Q3 al Máximo).
- Debe tenerse presente que cada uno de los intervalos tiene la misma cantidad de datos, por lo tanto los intervalos de menor longitud representan una mayor densidad. Es decir, los intervalos más cortos representan una mayor aglomeración de datos.



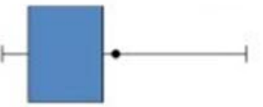
Veamos las formas que pueden tener los gráficos de caja y su interpretación:

| La mediana se encuentra justo en la mitad de la caja (Distribución Simétrica)       | La mediana se encuentra ubicada hacia la izquierda de la caja (Distribución Asimétrica hacia la izquierda) | La mediana se encuentra ubicada hacia la derecha de la caja (Distribución Asimétrica hacia la derecha) |
|---|--|--|
|  |                        |                   |
|  |                        |                   |
| Hay igual número de datos por encima que por debajo de la mediana.                  | Los Datos se concentran hacia los valores menores y los mayores están más dispersos.                       | Los datos se concentran hacia los valores mayores y los menores se están más dispersos.                |

La mediana puede inclusive coincidir con los cuartiles o con los límites de los bigotes. Esto sucede cuando se concentran muchos datos en un mismo punto. Podría ser este un caso particular de una distribución sesgada o el caso de una distribución muy homogénea.

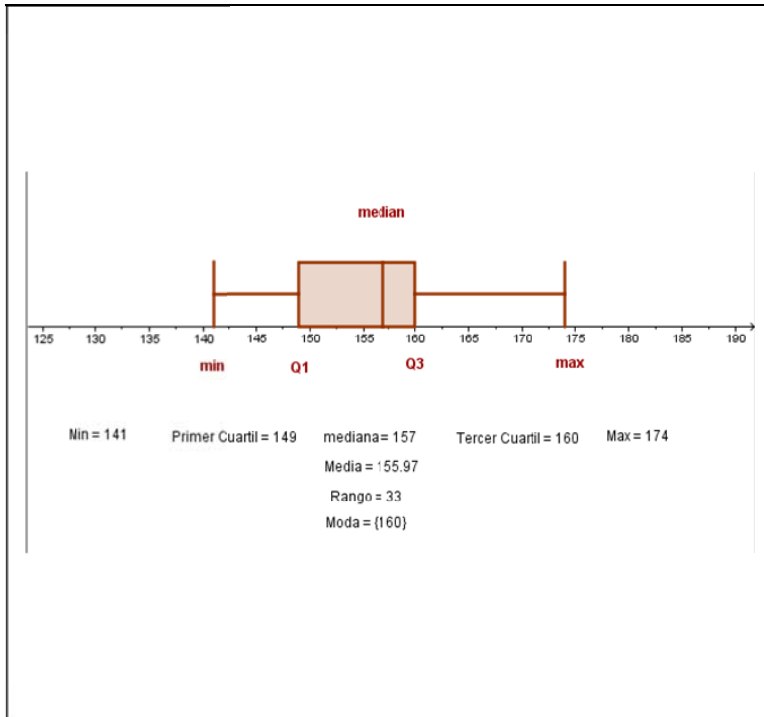


Otros casos para interpretación de gráficas

| Caso (A):  | Caso (B):   | Caso (C):  |
|--|---|--|
| Nos indica que la dispersión de datos es poca y no existen valores máximos y mínimos muy alejados de la mediana. Una caja de cuartiles pequeña indica que los datos son muy cercanos entre sí. | Respecto a la caja de cuartiles, se puede asegurar que este conjunto de datos posee una mayor dispersión, además de la posible existencia de valores atípicos (Alejados de la mediana) que alargan las líneas de máximo y mínimo. | Existen valores atípicos (Alejados de la mediana) que sesgaron el promedio y lo sacaron de la caja de cuartiles. |
|   |   |                             |

**APLICA LO APRENDIDO**

1. Las estaturas de un grupo de estudiantes se representaron en este diagrama de caja:

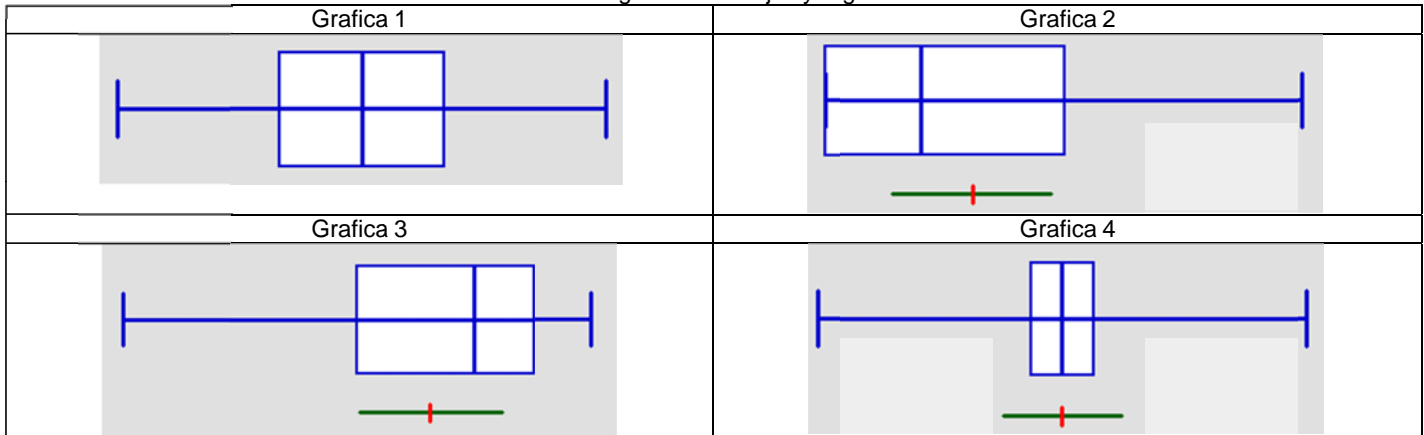


Con base en estos datos responda las siguientes preguntas:

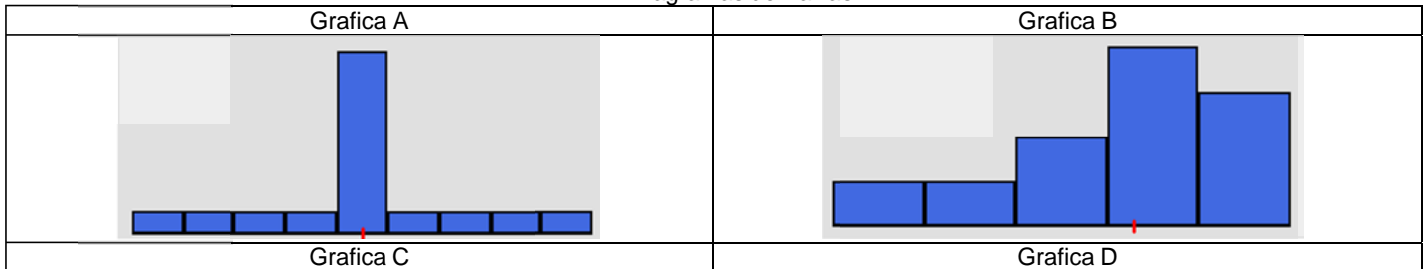
- Ordene los siguientes intervalos desde el más concentrado (1) hasta el más disperso (4)
  - Mínimo – Q1 ( )
  - Q1 – Mediana ( )
  - Mediana – Q3 ( )
  - Q3 – Máximo ( )
- Si el número de estudiantes que posee una estatura comprendida entre 149 y 157 es de 250. ¿Cuántos estudiantes fueron consultados?
- Si el número total de estudiantes consultados es de 450 ¿Cuántos estudiantes tiene una estatura comprendida entre?:
  - 149 a 157 \_\_\_\_\_
  - 141 a 160 \_\_\_\_\_
  - 149 a 174 \_\_\_\_\_
  - 149 a 160 \_\_\_\_\_
- Si el número de estudiantes que posee una estatura comprendida entre 150 y 155 es de 120. ¿Cuántos estudiantes fueron consultados?

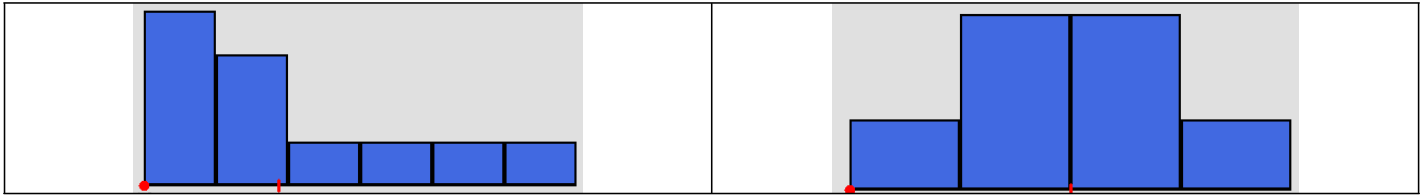
2. Dados los siguientes diagramas de cajas y bigotes y los siguientes diagramas de barras asigne a cada diagrama de cajas su correspondiente diagrama de barras. Justifique su respuesta

Diagramas de Cajas y Bigotes



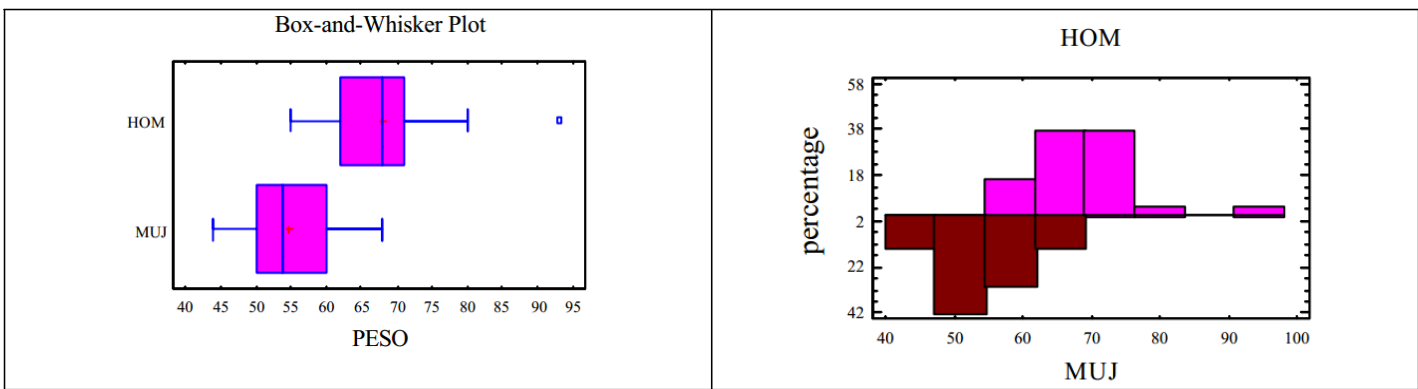
Diagramas de Barras





| Diagrama de Cajas y Bigotes | Diagramas de Barras | ¿Por qué? |
|-----------------------------|---------------------|-----------|
| 1                           |                     |           |
| 2                           |                     |           |
| 3                           |                     |           |
| 4                           |                     |           |

3. A continuación se muestran los histogramas y gráficos de caja correspondientes a la distribución de los pesos tanto de hombres como de mujeres



Existe algún dato extraño ¿Cuál es?

---



---

Cual grupo es más disperso? Los hombres o las mujeres?

---



---

La información del peso de los hombres se concentra hacia la izquierda o hacia la derecha?

---



---

La información del peso de las mujeres se concentra hacia la izquierda o hacia la derecha?

---



---

Con base en los histogramas y en los gráficos de caja Plantee por lo menos cuatro conclusiones.

---



---



---



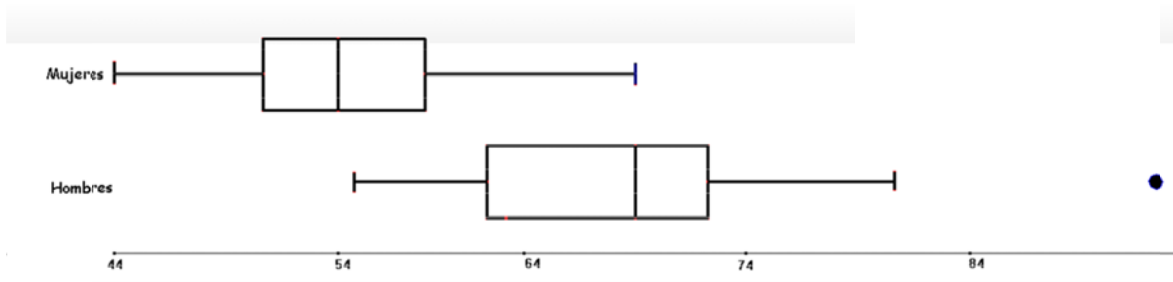
---

**IED TÉCNICO COMERCIAL MARIANO OSPINA RODRÍGUEZ - GUASCA**  
**IED GENERAL CARLOS ALBÁN - ALBÁN**  
**CUESTIONARIO DE ESTADÍSTICA**

**NOMBRE:** \_\_\_\_\_ **GRADO:** \_\_\_\_\_ **FECHA:** \_\_\_\_\_

**RESPONDA LAS PREGUNTAS 1 A 5 CON BASE EN LA SIGUIENTE SITUACIÓN**

Se ha tomado el peso (en kilogramos) a un grupo de 60 alumnos de bachillerato, diferenciando hombres y mujeres. Los resultados se muestran en su gráfico de cajas.



1. Identifique si cada una de las afirmaciones siguientes es verdadera o falsa. Justifique

a. Como el tamaño del gráfico de caja correspondiente a las mujeres es más grande, la cantidad de mujeres es mayor que la de los hombres. V \_\_\_\_ F \_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

b. La cantidad de hombres y mujeres debe ser aproximadamente igual ya que la extensión de los bigotes y de las cajas es similar en las dos gráficas. V \_\_\_\_ F \_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

2. ¿Cuál grupo presenta mayores pesos? Mujeres \_\_\_\_\_ Hombres \_\_\_\_\_ Justifique

\_\_\_\_\_

3. ¿Qué podría decir acerca de la distribución de los datos en ambos grupos respecto a:

a. ¿A la persona que pesa 54 Kg?

\_\_\_\_\_

\_\_\_\_\_

b. ¿A la persona que pesa 70 Kg?

\_\_\_\_\_

\_\_\_\_\_

4. a. ¿Qué porcentaje de los estudiantes de cada grupo pesa más de 50 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

\_\_\_\_\_

\_\_\_\_\_

b. ¿Qué porcentaje de los estudiantes de cada grupo pesa menos de 72 Kg? Justifique su respuesta describiendo el comportamiento del gráfico de cajas.

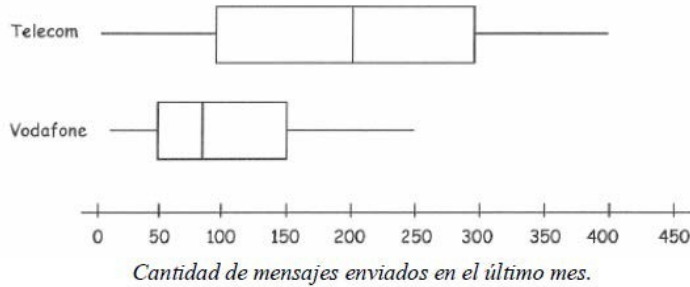


---

5. Identifique si hay o no datos atípicos en cada grupo, en caso afirmativo justifique porqué son atípicos

---

**RESPONDA LAS PREGUNTAS 6 A 10 Y GENERE UNA CONCLUSIÓN CON BASE EN LA SIGUIENTE SITUACIÓN:**  
A Paola le interesa hacer una comparación acerca de la cantidad de mensajes de texto que envían los usuarios de dos compañías de telefonía celular. Ella presentó la información a través de gráficos de caja como se muestra a continuación:



6. Paola quiere saber a qué compañía pertenecen los usuarios que más mensajes de texto enviaron durante el último mes ¿Qué le respondería y por qué?

---

7. a. ¿Entre qué par de valores está el 50% central de la cantidad de mensajes de texto en el último mes para cada una de las compañías? ¿Por qué?

---

b. Paola afirma que “Vodafone tiene una mayor dispersión en el número de mensajes que envían los usuarios que la que tiene Telecom”. Confirme o refute esta afirmación usando la información presentada en los gráficos de caja.

---

8. a. ¿Cuántos mensajes de texto envió el usuario que tuvo más consumo en cada una de las compañías? Justifique su respuesta.

---

b. ¿Cuántos mensajes de texto envió el usuario que menor consumo presentó en cada una de las compañías? Justifique su respuesta.

---

c. ¿Qué se podría concluir de esto al comparar los dos resultados anteriores?

---

9. a. Suponga que un nuevo usuario de Vodafone envió 450 mensajes de texto en un mes. ¿Este sería un dato atípico? Explique su respuesta

---

b. La inclusión de este nuevo dato afecta la forma del gráfico. Explique.

---

10. a. ¿En qué compañía los valores son más cercanos entre sí? \_\_\_\_\_.

b. ¿Qué podría decirse del 50% inferior y del 50% superior de los datos en esta compañía? \_\_\_\_\_.

---

---

---

---

---

**Considerando los gráficos de caja y las preguntas planteadas ¿qué compañía telefónica le aconsejaría escoger a Paola y por qué?**

---

---

---

