REPÚBLICA BOLIVARIANA DE VENEZUELA
UNIVERSIDAD DE LOS ANDES
FACULTAD DE HUMANIDADES Y EDUCACIÓN
ESCUELA DE EDUCACIÓN
DEPARTAMENTO DE MEDICIÓN Y EVALUACIÓN
MENCIÓN MATEMÁTICA

MÓDULO PARA LA ENSEÑANZA – APRENDIZAJE DE LA ESTADÍSTICA EN EL BACHILLERATO MEDIANTE LA SOLUCIÓN DE SITUACIONES PROBLEMAS

AUTORES:
JOAN FERNANDO CHIPIA LOBO
C.I. 17.662.609
CARMEN ZULEIMA LARA ANGEL
C.I. 12.992.614

TUTOR ACADÉMICO:
PROF. REINALDO CADENAS

MÉRIDA, 2008

UNIVERSIDAD DE LOS ANDES
FACULTAD DE HUMANIDADES Y EDUCACIÓN
ESCUELA DE EDUCACIÓN
DEPARTAMENTO DE MEDICIÓN Y EVALUACIÓN

ACTA DE GRADO

Hoy, veintiséis de septiembre de dos mil ocho, siendo las 10:30 pm, reunidos en el salón de Postgrado del Cindisi, Edif. B de la Facultad de Humanidades y Educación, los profesores: Reinaldo Cadenas (Tutor), Hermes Viloria (Jurado) y Gustavo Velasco (Jurado), designados para conocer la Memoria de Grado titulada: **"MÓDULO PARA LA ENSEÑANZA – APRENDIZAJE DE LA ESTADÍSTICA EN EL BACHILLERATO MEDIANTE LA SOLUCIÓN DE SITUACIONES PROBLEMAS"** presentada por los Bachilleres:

Joan F. Chipia L. y Carmen Z. Lara A.

Titulares de la Cédula de Identidad Nº V- **17.662.609 y 12.992.614**, respectivamente, en un todo de acuerdo a lo expuesto en el Artículo 25 del Reglamento de Memorias de Grado vigente y una vez cumplida la exposición pública del trabajo, este jurado decide calificarlos con:

VEINTE PUNTOS (20 PTOS)
CON MENCIÓN PUBLICACIÓN

*En consecuencia, los Bachilleres **Joan F. Chipia L.** y **Carmen Z. Lara A.**, han cumplido con todo los requisitos para optar al **Título de Licenciados en Educación Mención Matemática***

UNIVERSIDAD DE LOS ANDES
FACULTAD DE HUMANIDADES Y EDUCACION
DEPTO. DE MEDICION Y EVALUACION
MERIDA – VENEZUELA

Profesor Gustavo Velasco
Jurado

Profesor Hermes Viloria
Jurado

Cadenas

Profesor Reinaldo Cadenas
Tutor

DEDICATORIAS

Dedicatoria de Joan Fernando Chipia Lobo

A Dios Todopoderoso
A mis padres
A mis hermanos
A mis amigos

Dedicatoria de Carmen Zuleima Lara Angel

A Dios Todopoderoso
A mis padres
A mis hermanos y amigos
A mis compañeros
A mi hija Mariangel Nazareth

AGRADECIMIENTOS

A nuestro profesor y tutor académico ***Reinaldo Cadenas,*** por contribuir en nuestra formación como profesionales de la docencia, darnos su apoyo incondicional en la elaboración de esta investigación y habernos facilitado su valiosa e importante colaboración y conocimientos haciendo posible el logro de esta meta.

A la profesora ***Hazel Flores,*** por su asesoramiento y aportes de investigación que fueron fundamentales para la realización y culminación de esta memoria de grado.

A las Instituciones Educativas que favorecieron la realización de la fase diagnóstica base de nuestra investigación y a los docentes que prestaron su servicio para validar esta propuesta.

Al CDCHT por haber financiado la realización de este trabajo de investigación (Código H-1196-08-04-F) y de esta manera contribuir al crecimiento profesional de los estudiantes de pregrado.

A todas aquellas personas que de una u otra forma ayudaron en el alcance de esta meta.

Finalmente a todas las personas que estuvieron siempre en contra de esta investigación, muchas gracias porque cada día nos impulsaron más y más a la culminación de este trabajo.

TABLA DE CONTENIDOS	**Pág.**

LISTA DE TABLAS	Pág.

UNIVERSIDAD DE LOS ANDES
FACULTAD DE HUMANIDADES Y EDUCACIÓN
ESCUELA DE EDUCACIÓN
DEPARTAMENTO DE MEDICIÓN Y EVALUACIÓN

Título del Proyecto

Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato mediante la solución de situaciones problemas.

Autores:
Joan F. Chipia L. C.I.: V.-17.662.609
Carmen Z. Lara A. C.I.: V.-12.992.614

Tutor:
Reinaldo A. Cadenas A.

Fecha:
Octubre de 2008

Resumen

El presente trabajo tiene por objeto diseñar un Módulo que facilite el proceso de enseñanza – aprendizaje de las nociones básicas de Estadística a los docentes de la Tercera Etapa de Educación Básica y el segundo año del ciclo de Educación Media Diversificada, mediante el uso de la solución de situaciones problemas, donde el estudiante pueda construir activamente su aprendizaje. Para tal fin, se realizó una *Investigación Proyectiva*, ya que consiste en el diseño de un Módulo a través de una Propuesta de Orientación Didáctica para la enseñanza-aprendizaje de la Estadística en el bachillerato que permita superar las fallas detectadas en la revisión biblio - hemerográfica, curricular y diagnóstica para la construcción del aprendizaje de los conocimientos básicos de la Estadística y dar sentido a la problemática planteada en esta investigación. El proceso de investigación se llevó a cabo de la siguiente forma: se validó una prueba diagnóstica de conocimiento, luego se le aplicó a un grupo de estudiantes de diferentes instituciones de la ciudad de Mérida para obtener información sobre su preparación en Estadística y diagnosticar los conocimientos previos que poseen para atender las necesidades y dificultades de los alumnos en la enseñanza – aprendizaje de esta ciencia, esto se hizo con el fin de diseñar el Módulo de enseñanza – aprendizaje de la Estadística a través de una Propuesta de Orientación Didáctica como modelo didáctico, pedagógico, de planificación y desarrollo de las clases. Para validar la propuesta se seleccionan al azar cinco (5) docentes especialistas en el área de Matemática que impartan clases en el bachillerato y de este modo establecer la factibilidad de ser aplicada.

Palabras clave: Estadística, módulo, propuesta de orientación didáctica, situaciones problema, enseñanza-aprendizaje.

INTRODUCCIÓN

La presente investigación tiene por esencia diseñar un Módulo a través de una Propuesta de Orientación Didáctica para la enseñanza – aprendizaje de la Estadística en el bachillerato, partiendo del hecho de que la Educación debe ser un proceso que materialice una serie de habilidades, conocimientos, actitudes y valores adquiridos, produciendo cambios de carácter social, intelectual y emocional en la persona y logrando que éstos sean para toda su vida o por un periodo determinado.

La parte de la Educación que se desarrollará en la enseñanza - aprendizaje de la Matemática es la Estadística, que coloquialmente cuando se habla de ella, se suele pensar en una relación de datos numéricos presentada de forma ordenada y sistemática. Esta idea es la consecuencia del concepto popular que existe sobre el término y que cada vez está más extendido debido a la influencia de nuestro entorno, porque hoy día cualquier medio de comunicación (periódico, radio, televisión, etc.) aborda diariamente cualquier tipo de información Estadística sobre accidentes de tráfico, índices de crecimiento de población, turismo, tendencias políticas, etc.

Sólo cuando nos adentramos en un mundo más específico como es el campo de la investigación de las Ciencias Sociales: Medicina, Biología, Psicología,... empezamos a percibir que la Estadística no sólo es algo más, sino que se convierte en la única herramienta que, hoy por hoy, permite dar luz y obtener resultados, y por tanto beneficios, en cualquier tipo de estudio. Se puede definir la Estadística como la ciencia que estudia: cómo debe emplearse la información y cómo dar una guía de acción en situaciones prácticas que entrañan incertidumbre.

Sobre las bases de las consideraciones anteriores, la presente investigación, tiene por objeto diseñar un Módulo que facilite el proceso de enseñanza – aprendizaje de las nociones básicas de Estadística a los docentes de la Tercera Etapa de Educación Básica y el segundo año del ciclo de Educación Media Diversificada, mediante el uso de la solución de situaciones problemas, donde el estudiante pueda construir activamente su aprendizaje y adquiera un sentido personal, trascendental y de valor para el alumno.

Para tal fin, se realizará una *Investigación Proyectiva*, ya que consiste en el diseño de un Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato que permita superar las fallas detectadas en la revisión biblio-hemerográficas, curricular y diagnóstica para la construcción del aprendizaje de los conocimientos básicos de la Estadística y dar sentido a la problemática planteada inicialmente en esta investigación.

Esta investigación ha sido estructurada de la siguiente manera:

- En el capítulo 1 se realiza la descripción general del tema, la justificación de la investigación, el planteamiento del problema y los objetivos generales y específicos.

- En el capítulo 2 se presentan los antecedentes y los fundamentos teóricos que servirán de soporte para el desarrollo de la investigación.

- En el capítulo 3 se describe el marco metodológico para llevar a cabo la investigación.

- En el capítulo 4 se muestran los resultados del diagnóstico efectuado.

- En el capítulo 5 se presenta el Módulo para la enseñanza – aprendizaje de la Estadística mediante solución de situaciones problemas en el bachillerato (Municipio Libertador-Mérida) a través de una Propuesta de Orientación Didáctica.

- En el capítulo 6 se mostrará el análisis de los resultados obtenidos con respecto a la factibilidad del modelo propuesto, control y evaluación de procesos, conclusiones, limitaciones y recomendaciones finales.

CAPÍTULO 1. Contextualización de la temática

1.1.- *Descripción general del tema*

La Estadística es una ciencia intensamente dinámica y es un tema del programa básico del bachillerato venezolano, es de importancia vital en el proceso del aprendizaje escolar, tanto por su contribución al desarrollo cognitivo del adolescente, como por la funcionalidad que poseen la mayoría de los aprendizajes matemáticos en la vida adulta. Por otra parte, el área de Matemáticas en el bachillerato abarca dentro de sus objetivos, estudiar nociones elementales de Estadística Descriptiva (es la parte de la Estadística que describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos) y de Estadística Inferencial (es la parte de la Estadística en la cual se deducen propiedades o características de una población a partir de una muestra significativa) (Batanero, 2001).

La Estadística Descriptiva e Inferencial presentan dificultades en su proceso de enseñanza – aprendizaje, debido a que los alumnos no diferencian los tipos de variables y frecuencias y el cálculo e interpretación de las medidas de tendencia central (ver capítulo 4), por ésta razón, el objetivo fundamental de la presente investigación es diseñar un Módulo para la enseñanza - aprendizaje de la Estadística a través de situaciones problemas que permita desarrollar un modelo que propicie experiencias significativas en los estudiantes que transitan por esta etapa del bachillerato.

1.2.- *Justificación de la investigación*

El estudio de la Estadística permite al joven comprender situaciones de su entorno, tener un criterio para la toma de decisiones vinculadas a su ambiente escolar y familiar. Además, es un excelente medio para interrelacionar la Matemática con diversas áreas científicas y sociales, y para reforzar valores como la honestidad en la presentación de los resultados (Constitución Nacional de la República Bolivariana de Venezuela, 1999). En Venezuela la Estadística está incluida en el plan de estudio de la Tercera Etapa de Educación Básica y en el segundo año de Educación Media Diversificada e inclusive la mayoría de los programas universitarios incluye por lo menos un curso sobre Estadística. Pero, pese a esta presión social según Behar y Grima (2001), la preparación en la

Estadística no es suficientemente amplia pues la enseñanza-aprendizaje de esta disciplina ha sido relegada a un segundo plano.

Con el fin de no contribuir a que ésta ciencia sea relegada a un segundo plano, se debe buscar en el bachillerato que los estudiantes se enfrenten al aprendizaje de la Estadística de otra manera, dejando de lado la memorización de procedimientos y conceptos. Por ello, la investigación está orientada a propiciar una mejor comprensión y asimilación de los conceptos y métodos estadísticos. Esto ayudará a que los alumnos obtengan un mejor desenvolvimiento en su entorno social, además posean la capacidad de interpretar y evaluar la información Estadística, comunicando sus opiniones, criterios, análisis e interpretación acerca de esas informaciones. Para esto es necesario diseñar un Módulo de enseñanza – aprendizaje a través de situaciones problemas, para que los alumnos se sientan muy vinculados con lo que se está estudiando y con el estimulo de ser los protagonistas de su aprendizaje, dando paso a la comprensión e internalización de conocimientos estadísticos adecuados (Batanero, 2001).

Significa entonces que la importancia del módulo se centra en el estudiante siendo por esto novedosa y diferente a las propuestas de enseñanza – aprendizajes presentados anteriormente sobre Estadística en el bachillerato, la cual tendrá utilidad para los docentes y así éstos les muestren a los educandos ¿qué herramientas les proporciona la Estadística? A la vista de un conjunto de datos, ¿cómo manejarlos?, ¿cómo obtener conclusiones? Ante las “tradicionales” preguntas *¿cuándo estoy ante una situación real?, ¿cómo la resuelvo?,* y resuelvan la apatía de los alumnos que no encuentran ninguna conexión entre lo explicado en clase y la realidad. El módulo se presenta en clases bien diseñadas y encadenadas dejando de lado la memorización de procedimientos y conceptos para que el joven pueda resolver situaciones problemas, además de aplicar la Estadística porque ha sido comprendida mediante la utilización de recursos didácticos adecuados.

1.3.- *Planteamiento del problema*

La Estadística no es sólo una colección de conceptos y técnicas, también es una forma de razonar fomentando un razonamiento crítico, basado en la valoración de la evidencia objetiva. De este modo se adquiere la capacidad de lectura e interpretación de tablas y/o gráficos estadísticos que con frecuencia aparecen en los medios informativos y ayudan a comprender los temas del currículo de otras asignaturas. Esto se debe a su naturaleza interdisciplinar, donde los conceptos estadísticos aparecen en otras disciplinas, como en Ciencias Sociales, Biología, Geografía, Física, etc., lo que puede ocasionar conflictos cuando las definiciones o propiedades presentadas de los conceptos no son impartidas en las clases de Matemática (ver capítulo 4). Por su parte, Batanero y Godino (2001) resaltan que la importancia de esta ciencia en la actualidad se origina por la influencia de una civilización dominada por la ciencia y la tecnología, en una sociedad caracterizada por el cúmulo de información y la necesidad de tomar decisiones en el entorno que rodea al ser humano para desenvolverse en actividades que van desde la conducción de un país hasta actividades de esparcimiento y recreación; por ejemplo, si se conoce la cantidad de personas que usarán los servicios de un hotel en determinados momentos del año éste se puede organizar para su uso en ciertas temporadas, además a través de encuestas se puede lograr predecir el resultado de procesos electorales.

En efecto, la Estadística es una ciencia intensamente dinámica que permite predecir resultados y tomar decisiones de manera adecuada en esta sociedad caracterizada por el cúmulo de información. Como el área de Matemática del bachillerato abarca dentro de sus objetivos, estudiar nociones elementales de Estadística Descriptiva e Inferencial se debe presentar un alternativa distinta e innovadora para que los profesores se sientan motivados a enseñar esta ciencia en este nivel de estudio (Batanero, 2001). Así los estudiantes obtendrían un aprendizaje significativo considerando situaciones relacionadas con la comunidad y la institución y de este modo se contribuye a darle la transcendencia que posee esta ciencia, por lo tanto, para determinar los problemas acontecidos sobre la enseñanza-aprendizaje de la Estadística se hizo necesario realizar un diagnóstico.

En el diagnóstico efectuado, se revisaron en tres (3) instituciones educativas los planes de lapso de los tres (3) últimos años escolares donde solo en un año escolar de una de las instituciones, fue incluido la enseñanza de la Estadística lo cual evidenció que éste contenido no está siendo enseñado. También se entrevistaron (ver Anexo A) a tres (3) docentes especialistas en Matemática, arrojando que el 100% de los entrevistados no imparten la Estadística (Descriptiva e Inferencial) porque consideran más importantes otros contenidos del programa, están al final del currículo de estudio y no les alcanza el tiempo del año escolar. Esto trae como consecuencia que los alumnos no manejan los conceptos básicos de la Estadística (población, muestra y dato) ya que el 80% no respondió correctamente. Tampoco lo básico sobre la Estadística Descriptiva, porque no diferencian los tipos de variable debido a que 28 de los 30 evaluados respondió incorrectamente, el 76% de los estudiantes no diferencian los tipos de frecuencia y el 83% no lograron calcular e interpretar las medidas de tendencia central; además, no resuelven problemas donde se aplique la Estadística(ver capítulo 4), por ello, la investigación está orientada al diseño de un Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato propiciando una mejor comprensión y asimilación de los conceptos y métodos estadísticos en el Municipio Libertador del Estado Mérida.

1.4.- *Objetivos de la investigación*

1.4.1.- *Objetivo General*

Diseñar un Módulo que contribuya a mejorar el proceso de enseñanza – aprendizaje de la Estadística en los alumnos del bachillerato mediante la solución de situaciones problemas.

1.4.2.- *Objetivos Específicos*

- Identificar los conocimientos previos que deberían tener los alumnos para poder comprobar las necesidades y dificultades de la enseñanza – aprendizaje de la Estadística en el bachillerato.
- Diseñar el Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato.
- Validar el Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato.

CAPÍTULO 2. Marco teórico

2.1.- *Antecedentes*

Un estudio relevante es el de Moreno y Vallecillos (2002), este estudio tuvo por objeto manifestar que la enseñanza de la Estadística debería promover el uso del razonamiento estadístico en situaciones de la vida real, porque existe mayor vinculación del estudiantado hacia el contenido a obtener y esto les permitirá construir modelos de solución para los sucesos. También explicaron que enseñar Estadística implica conocer las nociones básicas de esta ciencia debido a que esto ayuda a interpretar y analizar de manera precisa los datos, en lugar de contentarse con explicaciones espontáneas y demasiado superficiales, como la repetición de contenidos de manera monótona. Este estudio concluyó que para mejorar la calidad de los aprendizajes estadísticos es necesario que las situaciones de la vida diaria estén inmersas en el sistema educativo, en un lugar y espacio determinado para que los estudiantes interpreten los resultados, utilicen la información y obtengan un aprendizaje significativo. Por lo tanto, como es necesario que las situaciones de la vida real estén inmersas en el proceso de enseñanza para lograr un aprendizaje significativo, se puede proyectar la realización de un módulo para la enseñanza-aprendizaje de esta ciencia mediante la solución de situaciones problemas.

Es conveniente mencionar la investigación realizada por Devia y Mora (2007) en Mérida, esta investigación tuvo como objetivo proveer al docente de herramientas útiles y alternativas en el proceso orientador sobre la enseñanza de las nociones elementales de Estadística Descriptiva mediante el uso del software SPSS como recurso de orientación didáctica. En este estudio se destaca la importancia de la solución de problemas y utilización de un software en el proceso enseñanza – aprendizaje de la Matemática, dado que los alumnos están, en general, profundamente atraídos por las nuevas tecnologías y situaciones de su entorno. El manejo de un software es un recurso que motiva el aprendizaje de la Estadística, dejando de lado la repetición mecánica e inconsciente de contenidos, además de permitir en el alumno el razonamiento constante, la comprensión y explicación de lo que está haciendo, este hecho representa, tanto para el joven como para el docente un inestimable espacio de aprendizaje que permite desarrollar y ejercitar las capacidades que hacen del ser humano un ser racional e inteligente. Los autores concluyen que los estudiantes de bachillerato consideran la Estadística como un tema sin importancia, complejo y difícil, por ello, el uso de un software incentiva el aprendizaje de las nociones elementales de la Estadística. Como el software es un recurso que motiva, se puede plantear la utilización de esta herramienta a manera de apoyo en el desarrollo del módulo para que favorezcan la comprensión e internalización de conocimientos estadísticos.

Otro estudio importante es el realizado por Batanero y Godino (2001), esta investigación tuvo dentro de sus objetivos analizar obstáculos de la Didáctica de la Estadística. Este estudio se interesó por explicar por qué el alumno se equivoca cuando se le pide realizar ciertas tareas, estas dificultades no se presentan de un modo aleatorio, ni imprevisible; con frecuencia se encuentran dificultades que se repiten, o que se producen regularmente en las tareas propuestas. Como el obstáculo es un conocimiento, el alumno utiliza este conocimiento para producir respuestas adaptadas a un cierto contexto que encuentra con frecuencia y cuando este conocimiento se usa fuera de este contexto genera respuestas incorrectas. Esta investigación concluyó que el alumno no es consciente del obstáculo y no logra establecer un conocimiento mejor, por tanto, es indispensable que el alumno se concientice del obstáculo, para adquirir un conocimiento más amplio. Este estudio es sumamente importante para la actual investigación porque se pueden plantear dentro del módulo tareas o problemas de consolidación que permitan a los alumnos superar los obstáculos y lograr un mejor conocimiento.

Un estudio que abarca el tema de los elementos que giran en torno a la resolución de problemas, fue el elaborado por Londoño (1995), con el objetivo de proveer al docente herramientas útiles para su proceso orientador. La solución de un problema no debe ser mecánica ni automática, por el contrario debe permitir en el estudiante el razonamiento constante, la comprensión y explicación de lo que está haciendo para que los alumnos se apropien del contenido logrando que el aprendizaje de la Matemática se convierta en un artículo de primera necesidad en su progreso correcto e indiscutible de manejar herramientas en la prosecución de su escolaridad. Este estudio concluye la importancia de la resolución de problemas en el proceso enseñanza – aprendizaje de la Matemática debido a que este hecho representa, tanto para el joven como para el docente, un inestimable espacio de aprendizaje que permite desarrollar y ejercitar las capacidades que hacen del ser humano un ser racional e inteligente. El estudio de Londoño denota una relación con la presente investigación, ya que ambas revelan la importancia de la resolución de problemas en el proceso de enseñanza-aprendizaje de un contenido matemático.

Finalmente, una investigación resaltante fue realizada por Medina (2006) en Tovar – Mérida, que tuvo por objetivo desarrollar un software educativo “Razzest 3” versión 1.0 como herramienta de aprendizaje del razonamiento estadístico, para mejorar el rendimiento académico en el área de Matemática en los alumnos de séptimo grado de Educación Básica. La reseñada investigación muestra la gran importancia que reviste el conocimiento básico sobre la Estadística y la resolución de problemas estadísticos a través de un software educativo; considerando la Estadística como una instrumento más en el proceso formativo que nos prepara para la vida en sociedad y poder generar riquezas desde el punto de vista social y humano. Sin embargo, una de las tendencias generales más difundidas en la actualidad, es dar a conocer los contenidos procedimentales a los alumnos a través de clases donde el joven actúa como un receptor pasivo que no motivan al educando a interesarse por el mundo de la Matemática. Esta investigación concluyó la importancia de emplear un software educativo para la enseñanza – aprendizaje de la Estadística, ya que en ese estudio se determinó que producto de la utilización de esta herramienta, aumentó el rendimiento académico en el área de Matemática, aunado a una mejor comprensión del tema. De este modo, se plantea la elaboración de un módulo que emplee la tecnología como herramienta de aprendizaje del razonamiento estadístico, permitiendo aumentar la motivación, mejorar el conocimiento del tema y lograr que el educando se convierta en un ente activo en su aprendizaje.

2.2.- *Bases teóricas*

2.2.1.- *Marco epistemológico*

El vocablo *statistik* proviene de la palabra italiana *statista* (que significa “estadista”). Fue utilizada por primera vez por Gouttfried Achenwall (1719 – 1772). El Dr. Zimmerman introdujo el término statistics (estadística) a Inglaterra. Su uso fue popularizado por sir John Sinclair. Sin embargo, mucho antes del siglo XVIII, la gente ya utilizaba y registraba datos (Levin y otros, 2004).

La Estadística es tan vieja como la historia registrada. Desde los comienzos de la civilización han existido formas naturales de Estadística, pues ya se usaban formas gráficas y otros símbolos en diversos materiales como pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas. Hacia el año 3000 a.C. los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola. Los egipcios analizaban los datos de la población mucho antes de construir las pirámides en el siglo XXXI a.C. Los libros bíblicos de Números y Crónicas incluyen, en algunas partes, trabajos de Estadística. En China existían registros numéricos similares con anterioridad al año 2000 a.C. Los griegos clásicos realizaban censos cuya información se utilizaba hacia el año 594 a.C. para cobrar impuestos. Precisamente fue un censo lo que motivó el viaje de José y María a Belén, según el Evangelio, pues el Imperio romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. Sin embargo sólo muy recientemente que la Estadística ha adquirido la categoría de ciencia (Da Silva, S/F).

En el siglo XVII surge la aritmética política, desde la escuela alemana de Conring, quien imparte un curso son este título en la Universidad de Helmsted. Posteriormente su discípulo Achenwall orienta su trabajo a la recogida y análisis de datos numéricos, con fines específicos y en base a los cuales se hacen estimaciones y conjeturas, es decir se observa ya los elementos básicos del método estadístico (Da Silva, S/F).

Para los aritméticos políticos de los siglos XVII y XVIII, la Estadística era el arte de gobernar; su función era la de servir de ojos y oídos al gobierno. La proliferación de tablas numéricas permitió observar la frecuencia de distintos sucesos y el descubrimiento de leyes estadísticas. Son ejemplos notables los estudios de Graunt sobre tablas de mortalidad y esperanza de vida a partir de los registros estadísticos de Londres desde 1592 a 1603 o los de Halley entre 1687 y 1691, para resolver el problema de las rentas vitalicias en las compañías de seguros (Da Silva, S/F).

En el siglo XIX aparecen las leyes de los grandes números con Bernouilli y Poisson. Otro problema que recibe gran interés por parte de los matemáticos de su tiempo, como Euler, Simpson, Lagrange, Laplace, Legendre y Gauss es el del ajuste de curvas a los datos. La Estadística logra con estos descubrimientos una relevancia científica creciente, siendo reconocida por la British Association for the Advancement of Science, como una sección en 1834, naciendo así la Royal Statistical Society. En el momento de su fundación se definió la Estadística como conjunto de hechos, en relación con el hombre, susceptibles de ser expresados en números, y lo suficiente numerosos para ser representados por leyes (Da Silva, S/F).

Se crearon poco a poco sociedades estadísticas y oficinas estadísticas para organizar la recogida de datos estadísticos; la primera de ellas en Francia en 1800. Como consecuencia, fue posible comparar las estadísticas de cada país en relación con los demás, para determinar los factores determinantes del crecimiento económico y comenzaron los congresos internacionales, con el fin de homogeneizar los métodos usados. El primero de ellos fue organizado por Quetelet en Bruselas en 1853 (Da Silva, S/F).

En nuestros días, la Estadística se ha convertido en un procedimiento efectivo para describir con precisión los valores de datos económicos, políticos, sociales, psicológicos, biológicos, físicos entre otros, y sirve de herramienta para relacionar y analizar dichos datos. La Estadística no consiste sólo en reunir y tabular datos, sino sobre todo en el proceso de interpretación de esa información. Además el desarrollo de la teoría de la Probabilidad ha aumentado el alcance de las aplicaciones de la Estadística. Muchos conjuntos de datos se pueden aproximar, con gran exactitud, utilizando determinadas distribuciones probabilísticas y los resultados de éstas se pueden utilizar para realizar un análisis más exhaustivo y confiable (Batanero, 2001).

2.2.2.- *Marco psicopedagógico*

El concepto de problema, en el sentido piagetiano, se entiende como una situación de desequilibrio que “afecta” las estructuras intelectuales (Piaget y Beth, 1980). Solucionar esta condición o resolver el problema equivale, en este mismo sentido, al logro de un nuevo estado de equilibrio. Este nuevo estado implica una estructura más compleja, que hará posible la realización, “no problematizada”, de la actividad propuesta inicialmente como situación problemática.

Por su parte, Parra (1995), define un problema matemático como una situación no resuelta, donde la matemática presenta un papel fundamental para la solución del mismo. Así, por ejemplo, cuando se plantea a un niño que conoce el proceso de la operación adición, la realización de una suma con número enteros positivos, responderá "automáticamente" a:

¿Cuál es el resultado de la adición de 12 + 5?

Pero si a ese niño le preguntamos inmediatamente:

¿Cuál es el resultado de la adición de – 12+ 5?

La respuesta dejará de ser automática, pues habrá que esperar que el niño construya los esquemas mentales que le permitan la internalización del desconocido conjunto numérico de los Enteros (Z) y comportamiento de los números negativos. Con el ejemplo anterior se pone en evidencia que, en la primera pregunta presentada al niño, no existe un problema en el sentido estricto de la palabra, puesto que en sus estructuras mentales, disponibles para ser activadas ante situaciones como esta, existen los elementos suficientes para realizar la asimilación o comprensión de los elementos involucrados en la pregunta: números positivos y operación adición (Parra, 1995). En la segunda pregunta, la operación adición pierde su sentido cuando se introduce un nuevo elemento a manejar, cuando se necesita conocer el comportamiento de los números negativos, no conocidos por el niño. En este caso la situación es de problema, de conflicto, de no tener los elementos indispensables para dar una respuesta inmediata y eficaz al cuestionario enfrentado. Para que continúe el proceso del desarrollo cognoscitivo, lógico – matemático es necesario, entonces, que mediante un proceso pedagógico determinado, se permita al niño alcanzar a comprender el sentido de los números negativos (Londoño, 1995). El niño podrá incorporar estos nuevos números a su estructura intelectual, que se amoldará y podrá responder, después de algunas situaciones de práctica, tan automáticamente a la segunda pregunta como a la primera.

Parra (1995) estableció cinco criterios básicos para caracterizar un buen problema: el primer criterio que se debe tomar en cuenta para saber si un problema es o no favorable, es la redacción, el problema deberá estar bien estructurado, de manera que sea entendible. Ejemplo: Observe la siguiente serie, según la relación entre los números. ¿Cuál es el que le sigue al último número? 3, 11, 19, ___. Un segundo criterio es la familiaridad del enunciado, un problema puede estar muy bien redactado, pero si éste no es familiar a las experiencias previas de los alumnos nunca podrá ser comprendido. Ejemplo: Clasificar las siguientes variables: marcas de cerveza, nivel educativo (primario secundario, superior), el peso en kilogramos, La temperatura de un enfermo en grados celsius. Otro nivel es el de razonamiento, un problema no tiene una solución mecánica, directa, porque si así fuese, la resolución del problema perdiese su esencia misma, convirtiéndose en una rutina mecánica. Ejemplo: determine de las Medidas de Tendencia Central ¿Cuál es la más adecuada para resumir los datos del ingreso económico en bolívares cuando existe un valor extremo? Un cuarto criterio es que debe tener suficiencia de elementos, esto es, datos, condición y una situación no resuelta que no necesariamente tiene que estar planteada en forma de pregunta, y que se denomina incógnita. Ejemplo: Si 56 + n = 81 ¿Qué número es n? Asimismo, un problema debe estar adaptado al nivel del alumno, ya que se parte del deber de respetar las capacidades intelectuales del mismo. Por último y no por esto menos importante, un problema debe ser original, de manera que fije en el alumno desde un principio un desafió, el de resolverlo.

Behar y Grima (2001) señalaron que la manera más original de presentar un problema es a través de una *situación problema:* es el planteamiento de un conjunto de factores o circunstancias que afectan a alguien o algo en un determinado momento y cuya respuesta desconocida debe obtenerse a través de métodos adecuados, tales como el método científico o métodos estadísticos.

¿Qué es una propuesta? Es un documento que describe el proyecto de investigación científica, tecnológica o de innovación que pretenden ejecutar instituciones, universidades, centros, laboratorios, empresas y demás personas que se encuentren involucrados a nivel individual o como grupo, contribuyendo a identificar el problema o necesidad y conseguir las metas propuestas (Sin autor, 2002).

¿Qué es un módulo educativo? Es un material educativo que agiliza y fortalece el aprendizaje de los estudiantes en el aula, desarrollando y modificando los factores que intervienen en el proceso de enseñanza-aprendizaje para obtener resultados diferentes (Barrera, 2004).

¿Para qué desarrollar un módulo educativo? Porque permite al docente una serie de situaciones de aprendizaje diferentes adecuándose a una situación específica de

aprendizaje. También se debe destacar el hecho que mediante estos materiales educativos la creatividad del alumno es fundamental debido a que no tienen que enfrentarse a un texto rígido que le muestra que debe hacer exactamente, sino más bien puede sugerir y adecuar las actividades de acuerdo a sus expectativas, curiosidad y posibilidades, de ahí se rescata que los módulos de educativos deberían de promover esto por la contextualización de los aprendizajes de acuerdo a cada región en la que estudiante vaya a utilizar esta herramienta (Barrera, 2004).

¿Qué es una propuesta de orientación didáctica? Es un documento o módulo que recoge las aportaciones del trabajo de campo, y procura dar respuesta a las informaciones obtenidas y a las reflexiones críticas realizadas. Contiene una serie de principios generales: orientaciones para desarrollar una didáctica de investigación, pautas para llevar a la práctica, modelos para trabajar cooperativamente en el aula; orientada hacia la renovación de la práctica docente y centrada en el modelo del profesor investigador (Batanero y Godino, 2001).

¿Por qué desarrollar un módulo a través de una propuesta de orientación didáctica para la Estadística? Porque es una alternativa innovadora, motivadora y diferente a lo presentado en los libros de texto y las propuestas presentadas por los docentes y esto permitirá que los alumnos posean la capacidad para interpretar y evaluar críticamente la información Estadística y sus argumentos estén apoyados en datos que las personas pueden encontrar en diversos contextos, incluyendo los medios de comunicación, pero no limitándose a ellos y la capacidad para discutir o comunicar sus opiniones respecto a tales informaciones estadísticas cuando sea relevante. Así mismo señala Barrera (2004) la propuesta es preferible que se desarrolle bajo un enfoque constructivista, debido a que es una pedagogía centrada en el alumno, donde el proceso de enseñanza-aprendizaje se apoya en sus necesidades, intereses y experiencias, reconociendo las características individuales y culturales, situando los aprendizaje dentro de ese contexto. Entonces la propuesta está basada en la asistencia y apoyo, donde el docente debe ser guía y orientador de los procesos de aprendizaje, en este sentido el módulo debe ser una herramienta para lograr que el estudiante construya su aprendizaje y el docente lo debe conocer y aplicar en beneficio de sus alumnos.

El objetivo de este Módulo educativo no es únicamente enseñarle al alumno una serie de métodos estadísticos, sino que además construya, comprenda e interprete las nociones desarrolladas en el bachillerato de Estadística; además, son dos los fines fundamentales de la enseñanza de esta ciencia en la escuela: que los alumnos lleguen a comprender y a apreciar el papel de la Estadística en la sociedad, incluyendo sus diferentes campos de aplicación y el modo en que la Estadística ha contribuido a su desarrollo y que los alumnos lleguen a comprender y a valorar el método estadístico, esto es, la clase de preguntas que un uso inteligente de la Estadística puede responder, las formas básicas de razonamiento estadístico, su potencia y limitaciones.

2.2.3.- *Marco matemático*

Se presentan las definiciones de algunas nociones elementales y básicas, que permitan lograr una comprensión real de lo que es la Estadística, induciendo al estudiante en el uso y manejo de datos numéricos contextualizados, para que distinga y clasifique los conceptos:

Estadística: es la ciencia que se encarga de la recopilación, organización, presentación, resumen y análisis de una serie de observaciones o informaciones en un problema planteado de una investigación. Una vez resumida y analizada la información puede ser empleada por el investigador para tomar decisiones, hacer predicciones y estimaciones que están sometidas a cierto riesgo o incertidumbre (Torres, 1997). Esta ciencia se divide en: *Estadística Descriptiva:* es la parte de la Estadística que tiene como objetivo la recopilación, presentación y organización de un conjunto de datos, con la finalidad de describir en la forma más apropiada la información arrojada por éstos (Pestaña, 2002). *Estadística Inferencial:* es la parte de la Estadística que tiene como objetivo estimar las características de una población, a partir de datos referentes a la muestra (Pestaña, 2002).

Población (N): es el conjunto de todos los elementos que cumplen ciertas propiedades comunes y proporcionan información sobre un problema planteado en una investigación y sobre quienes se van hacer extensivas las conclusiones de la investigación (Torres, 1997). *Muestra (n):* es un subconjunto o parte de la población que posee las mismas características o rasgos de ésta (Torres, 1997). *Dato* $\left(x_i\right)$: son los registros de los resultados de una serie de medidas y/u observaciones de un fenómeno o evento determinado (Pestaña, 2002).

Variable estadística: son el conjunto de propiedades, rasgos o cualidades presentes en un problema de investigación, se obtienen a través de la muestra y los elementos de la población. Se puede dividir en *categorías:* es cada uno de los conjuntos básicos en los que puede clasificarse cada variable estadística. Los tipos de variable son: *variables cualitativas:* son aquellas variables estadísticas que están relacionadas con las características propias del conjunto de elementos de la muestra o población, como por ejemplo: estado civil, nacionalidad, religión y otras. V*ariables cuasi-cuantitativas:* son aquellas variables estadísticas que están relacionadas con las características específicas del conjunto de elementos de la muestra o población, con la peculiaridad de que se pueden ordenar entre las categorías de la misma, como por ejemplo: nivel socio – económico (alto, medio, bajo), rango militar y otros. Y las v*ariables cuantitativas:* estas variables estadísticas son caracterizadas por información cuantificable y numérica que se le asocia al conjunto de individuos de una población o muestra. Esta caracterización puede ser clasificada en *discreta*: son aquellas cuyo valor está determinado por valores enteros, como por ejemplo: número de hijos, número de hojas de un libro de Estadística y otras. C*ontinua*: son aquellas cuyo valor está determinado por valores reales, generalmente estas variables pueden tomar toda una gama de valores dentro de la recta real, como por ejemplos: estatura en metros (m.), peso en kilogramos (kg.), distancia en centímetros (cm.), tiempo en segundos (s.) y otras (Torres, 1997).

Representación y organización de datos: una vez comprendidos los conceptos básicos podemos realizar un estudio descriptivo de una colección de datos, realizando tablas o gráficos adecuados al tipo de variable.

Tabla: consiste en presentar los datos organizadamente en arreglos tabulares. La presentación tabular constituye la forma más general de presentar datos estadísticos (Armas, 1988). *Componentes o elementos básicos de una tabla*:

Género	ni	fi	%
Masculino	13	0,41	41
Femenino	19	0,59	59
Total	32	1	100

Título: se debe mencionar los sujetos a clasificar, la variable observada, la ubicación espacial y temporal, y la fecha. *Encabezado:* se debe identificar cada columna con un nombre o símbolo, correspondiente a la variable o a la frecuencia respectiva.

Cuerpo de la tabla: se identifican las modalidades, categorías o clases de cada una de las frecuencias y sus porcentajes de ser necesarios, por ello definimos las frecuencias y porcentajes: Frecuencia absoluta (n_i): es el número de veces que se repite un dato (x_i), se puede hallar para cualquier tipo de variable. Frecuencia absoluta acumulada (N_i): es la frecuencia de cada dato (x_i) más la suma de los valores anteriores a dicha suma, solo para variables cuasi–cuantitativas y cuantitativas. Se halla así:

$$N_i = n_1 + n_2 + \ldots + n_i = \sum_{j=1}^{i} n_j .$$

Frecuencia relativa simple (f_i): es una proporción del número de datos que se repiten entre el total de datos de la muestra, es decir, es el cociente de la frecuencia absoluta entre el total de datos de la muestra, se puede hallar para cualquier tipo de variable. Se calcula así:

$$f_i = \frac{n_i}{n}.$$

Frecuencia relativa acumulada (F_i): es una proporción del número de datos acumulados entre el total de datos de la muestra, es decir, el cociente de la frecuencia acumulada entre el total de la muestra, solo para variables cuasi–cuantitativas y cuantitativas. Se calcula así:

$$F_i = \frac{N_i}{n} = \frac{n_1 + n_2 + \ldots + n_i}{n} = f_1 + f_2 + \ldots + f_i = \sum_{j=1}^{i} f_i .$$

Porcentaje (%): es una forma de expresar una proporción o fracción como un cociente de denominador cien, es decir, el producto de la frecuencia relativa simple por cien, se puede hallar para cualquier tipo de variable. Se calcula así:

$$\% = \frac{n_i}{n} \times 100 = f_i \times 100.$$

Porcentaje acumulado (P.A.): es una forma de expresar una proporción o fracción acumulada como una fracción de denominador cien, es decir, el producto de la frecuencia relativa acumulada por cien, solo para variables cuasi – cuantitativas y cuantitativas. Se calcula así:

$$P.A. = \frac{N_i}{n} \times 100 = F_i \times 100.$$

<u>*Fuente:*</u> es el registro o archivo de donde se tomó la información, debiéndose especificar si la misma no se encuentra archivada, es un elemento indispensable en toda tabla ya que nos indica el origen de la información. Se coloca en la parte inferior del cuadro.

Distribución de frecuencias: este tipo de tablas se realiza para agrupar grandes cantidades de datos cuantitativos en clases y deben cumplir con que la clasificación de datos debe ser *mutuamente excluyente*, es decir, un valor no puede estar en dos clases a la vez y *colectivamente exhaustiva*, es decir, todos los sujetos deben estar clasificados, no debe quedar ninguno fuera de la clasificación o distribución (Armas, 1988). *Para la construcción de una distribución de frecuencias son necesarios los siguientes elementos: recorrido de la clase o intervalo* (R)*:* es la distancia o diferencia que existe entre el máximo y el mínimo valor de una serie de datos más la unidad. Se calcula así: $R = (Valor \quad máximo - \quad Valor \quad mínimo) + unidad$. *El número de clases o intervalos* (Nc)*:* es un valor arbitrario por convención que vamos a tomar de acuerdo al criterio del investigador, se recomienda que el número de clases no sea inferior a cinco (5) ni mayor a diez (10). Un número muy pequeño de clases significaría una agrupación extrema, y en consecuencia sin utilidad; y un número muy grande no nos estaría resolviendo el problema de resumir los datos. *Amplitud de la clase o intervalo* (a_i)*:* es la proporción del rango de la clase entre el número de clases. Se calcula así: $a_i = \frac{R}{Nc}$ (Armas, 1988).

Límites de clase: son los extremos de la clase o intervalo.

Límite aparente inferior $(L.A.I.)$*:* es el extremo inferior o valor mínimo del intervalo. Se calcula así: $L.A.I. = Valor \quad mínimo$.

Límite aparente superior $(L.A.S.)$*:* es el extremo superior o valor máximo del intervalo. Se calcula así: $L.A.S. = (Valor \quad mínimo + a_i) + unidad.$

Límite real inferior$(L.R.I.)$*:* $L.R.I. = L.A.I. - \left(\frac{unidad}{2}\right)$.

Límite real superior$(L.R.S.)$*:* $L.R.S. = L.A.S. + \left(\frac{unidad}{2}\right)$(Armas, 1988).

Nota: la unidad que se considera en las fórmulas presentadas dependen de cómo están presentados los datos, si todos los valores son números enteros la unidad es uno, por tanto va dependiendo de la unidad seguida de cero.

Gráficos: los gráficos trasmiten en forma inmediata una idea general sobre los principales aspectos de los datos pero no proporcionan detalles. Un gráfico debe ser sencillo, de fácil interpretación y sólo debe suministrar valores aproximados. Son muy variados los tipos de gráficos, y por este motivo se realiza una selección de los necesarios para desarrollar la representación gráfica en el bachillerato (Armas, 1988).

Gráfico de barras simples: se recomienda para datos cualitativos y cuasi – cuantitativos. Para construir este gráfico, se utiliza en el eje de las abscisas las categorías y en el eje de las ordenadas, el número de datos de la variable o frecuencia absoluta de cada categoría. Para cada categoría se levanta una barra en forma rectangular cuya altura viene dada por el número de datos de cada categoría. Las barras deben ir separadas y tanto el ancho como la distancia que las separa son arbitrarios, pero una vez fijados deben mantenerse en todo el gráfico (Armas, 1988).

Gráfico circular o de sectores: se recomienda para variables cualitativas y cuasi – cuantitativas. Se construye bajo la base de dividir el círculo en sectores (S_i), en tantos sectores como modalidades tenga la variable, de tal manera que cada sector sea proporcional al número de datos de la correspondiente modalidad. Es necesario determinar cuántos de los trescientos sesenta grados del círculo corresponden a cada sector, se calcula así: $S_i = n_i \times \frac{360}{n}$. (Armas, 1988).

Histograma: se utiliza para representar distribuciones de frecuencias cuyas clases son intervalos. Consiste en un gráfico de barras rectangulares con la particularidad de que las barras están juntas unas de otras y el ancho de la misma es arbitrario, pero una vez fijada deben mantenerse en todo el gráfico. Se construye llevando sobre el eje de abscisas los límites reales de las clases, luego en el eje de las ordenadas la frecuencia absoluta, luego se unen los puntos medios o marcas de clase (X_m) con su respectiva frecuencia absoluta, lo cual determina la atura de la barra.

Nota: La marca de clase o punto medio (X_m). Se calcula así: $X_m = \frac{L.A.I.+L.A.S.}{2} = \frac{L.R.I.+L.R.S.}{2}$.(Armas, 1988).

Medidas descriptivas numéricas: cuando se trate de una cantidad muy numerosa de datos, la condensación y la representación gráfica de los mismos, constituyen elementos muy importantes pero no suficientes para realizar un adecuado análisis descriptivo de una colección de ellos. Se hace necesario utilizar ciertos indicadores numéricos que proporcionan una idea clara sobre aspectos relevantes, estos indicadores son: las medidas de tendencia no central y central y las medidas de dispersión o de variabilidad, que se estudiarán con más detalle a continuación:

Medidas de tendencia no central: las medidas de posición no central permiten ubicar la posición que ocupa un valor dentro de un conjunto de datos, se calcula para variables de tipo cuasi-cuantitativo o cuantitativo y vienen expresadas en las mismas unidades en que vienen expresados los datos (Armas, 1988).

Percentiles(P_k)*:* son aquellos valores que dividen los datos ordenados en cien partes iguales. Existen noventa y nueve percentiles que se denotan por $P_k, donde\ k = \{1,2,\ldots,99\}$, y el subíndice se refiere al porcentaje de casos por debajo del percentil. Entre dos percentiles consecutivos cualesquiera se encuentra un 1% o 1/100 partes de los datos. Se calcula así para *datos sin agrupar*: $P_k = \frac{n \times k}{100}$.(Armas, 1988).

Para datos agrupados:

$$P_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{100}\right) - N_{i(a)}}{n_i} \right) \times a_i \ .$$

Siendo:

P_K : Percentil a calcular.

$L.R.I.$: Límite real inferior de la clase del percentil.

n : Número total de datos de la muestra.

$N_{i(a)}$: Frecuencia acumulada anterior a la clase del percentil.

n_i: Frecuencia absoluta de la clase del percentil.

a_i: Amplitud de la clase del percentil.

Deciles(D_k)*:* son aquellos valores que dividen los datos ordenados en diez partes iguales. Existen nueve que se denotan por$D_k, donde\ k = \{1,2,\ldots,9\}$, y el subíndice se refiere al porcentaje de casos por debajo del percentil. Entre dos percentiles consecutivos cualesquiera se encuentra un 10% o 1/10 partes de los datos. *Se calcula así para datos sin agrupar:* $D_k = \frac{n \times k}{10}$.(Armas, 1988).

Para datos agrupados:

$$D_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{10} \right) - N_{i(a)}}{n_i} \right) \times a_i \, .$$

Siendo:
D_K : Decil a calcular.
$L.R.I.$: Límite real inferior de la clase del decil.
n : Número total de datos de la muestra.
$N_{i(a)}$: Frecuencia acumulada anterior a la clase del decil.
n_i: Frecuencia absoluta de la clase del decil.
a_i: Amplitud de la clase del decil.

Cuartiles(Q_k)*:* son aquellos valores que dividen los datos ordenados en cuatro partes iguales. Existen tres que se denotan por $Q_k, donde\ k = \{1,2,3\}$, y el subíndice se refiere al porcentaje de casos por debajo del percentil. Entre dos percentiles consecutivos cualesquiera se encuentra un 25% o 1/4 partes de los datos. Se calcula así para datos sin agrupar: $Q_k = \frac{n \times k}{4}$.(Armas, 1988).

Para datos agrupados:

$$Q_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{4} \right) - N_{i(a)}}{n_i} \right) \times a_i \, .$$

Siendo:
Q_K : Cuartil a calcular.
$L.R.I.$: Límite real inferior de la clase del cuartil.
n : Número total de datos de la muestra.
$N_{i(a)}$: Frecuencia acumulada anterior a la clase del cuartil.
n_i: Frecuencia absoluta de la clase del cuartil.
a_i: Amplitud de la clase del cuartil.

<u>Medidas de tendencia central:</u> se refieren al punto medio de una distribución de datos, también son conocidas como medidas de posición central y determinan hacia donde tienden a concentrarse las observaciones o alrededor de la cual se distribuyen el conjunto de datos y vienen expresada en las mismas unidades de los datos (Levin y otros, 2004).

Moda (Mo): de una serie de datos se define como aquel valor que se repite más o de mayor frecuencia. Es la única medida de posición que se puede hallar para los diferentes tipos de variables. En los datos es posible tener más de una moda e incluso es posible que no exista. Esto último ocurre cuando no hay un valor que se repita más que los demás. (Armas, 1988).

Para datos agrupados se calcula así:

$$Mo = L.R.I. + \left(\frac{d_1}{d_1 + d_2} \right) \times a_i.$$

Siendo:

$L.R.I.$: Límite real inferior de la clase modal.

a_i: Amplitud de la clase modal

$d_1 = Frecuencia\ modal - Frecuencia\ anterior$

$d_2 = Frecuencia\ moda - Frecuencia\ posterior$

Mediana (Md): es aquel valor que ocupa la posición central de los datos de una compilación de datos que previamente han sido ordenados en forma creciente o decreciente. Cuando el número de datos es impar, hay un solo valor que ocupa la posición central y en consecuencia ese dato es la mediana, mientras que cuando el número de datos es par, existen dos valores que ocupan la posición central entonces toma como mediana el promedio entre estos (Armas, 1988).

Para datos agrupados se calcula así:

$$Md = L.R.I + \left(\frac{n/2 - N_{i(a)}}{n_i} \right) \times a_i.$$

Siendo:

$L.R.I.$: Límite real inferior de la clase de la mediana.

n : Número total de datos de la muestra.

$N_{i(a)}$: Frecuencia acumulada anterior a la clase medianal.

n_i: Frecuencia absoluta de la clase medianal.

a_i: Amplitud de la clase medianal

Media aritmética $(\bar{x})$*:* es la suma de los datos dividida entre el número de sumandos de una colección de datos $x_1, x_2, \dots, x_{n-1}, x_n$. se interpreta como el valor promedio de una serie de datos y se calcula solo para variables cuantitativas (Armas, 1988). Se calcula así:

Para datos sin agrupar:

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

Para datos agrupados:

$$\bar{x} = \frac{X_{m_1} + \dots + X_{m_k}}{n_1 + \dots + n_k} = \frac{\sum_{i=1}^{k} X_{m_i} n_i}{\sum_{i=1}^{k} n_i}.$$

<u>*Medidas de dispersión o de variabilidad:*</u> son las que permiten cuantificar o expresar la variación de una distribución de datos. Para efecto de la investigación se utilizaran las Absolutas (permiten comparar las variables de dos o más distribuciones expresadas en la misma unidad de medida) (Levin y otros, 2004).

Rango (R)*:* es la diferencia entre los valores extremos de una distribución y es la medida de variabilidad más sencilla (Armas, 1988). Se calcula así: $R = (Valor \quad máximo - \quad Valor \quad mínimo)$.

Rango percentílico (RP)*:* es la diferencia entre el percentil noventa (P_{90}) y el percentil diez (P_{10}) por tanto es una medida de dispersión basada en los percentiles. Esta medida también puede expresarse equivalentemente como un intervalo de la forma $(P_{10}; P_{90})$. Este rango o intervalo incluye el 80% de los datos ubicados en la parte central de la distribución; es decir, excluye el 10% inferior de los datos y el 10% superior de los datos. Para datos no agrupados y agrupados se calcula con la siguiente formula: $RP = P_{90} - P_{10}$ (Armas, 1988).

Rango cuartílico (RQ)*:* es la diferencia entre el cuartil uno (Q_1) y el cuartil tres (Q_3), por tanto es una medida de dispersión basada en cuartiles. Esta medida también puede expresarse equivalentemente como un intervalo de la forma $(Q_1; Q_3)$. Este rango o intervalo incluye el 50% de los datos ubicados en la parte central de la distribución; es decir, excluye el 25% inferior de los datos y el 25% superior de los datos. Para datos no agrupados y agrupados se calcula con la siguiente formula: $RQ = Q_3 - Q_1$ (Armas, 1988).

Desviación media (DM)*:* es la media aritmética de los desvíos o diferencias en valor absoluto, de cada uno de los datos con respecto a su media (Armas, 1988). Se calcula así:

Para datos sin agrupar:

$$DM = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}.$$

Para datos agrupados:

$$DM = \frac{\sum_{i=1}^{k} \left| X_m - \bar{x} \right| \times n_i}{n}, donde \quad k = Nc.$$

Desviación mediana (DMd)*:* es el valor central de los desvíos o diferencias en valor absoluto, de cada uno de los datos con respecto a su mediana (Armas, 1988). Se calcula así:

Para datos sin agrupar:

$$DMd = \frac{\sum_{i=1}^{n}\left|x_i - \overline{\overline{x}}\right|}{n}.$$

Para datos agrupados:

$$DM = \frac{\sum_{i=1}^{k}\left|X_m - Md\right| \times n_i}{n}, donde \quad k = Nc.$$

Varianza muestral (S^2)*:* es la media aritmética de las desviaciones cuadráticas de los valores de la variable respecto a la media aritmética. Cuando los datos corresponden a un grupo o muestra se toman en cuenta $n-1$ datos. (Armas, 1988). Se calcula así

Para datos no agrupados

$$S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}.$$

Para datos agrupados:

$$S^2 = \frac{\sum\left(X_m - \bar{x}\right)^2 \times n_i}{n-1}.$$

Varianza poblacional (σ^2)*:* es la media aritmética de las desviaciones cuadráticas de los valores de la variable respecto a la media aritmética; toma en cuenta todos los sujetos de la población y la estimamos a partir de la varianza muestral (Armas, 1988). Se calcula así:

Para datos sin agrupar:

$$\sigma^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n}.$$

Para datos agrupados:

$$\sigma^2 = \frac{\sum\left(X_m - \bar{x}\right)^2 \times n_i}{n}.$$

Desviación típica o estándar muestral (S)*:* es la raíz cuadrada de la media aritmética de las desviaciones cuadráticas de los valores de la variable, con respecto a la media aritmética (Armas, 1988).

Desviación típica o estándar poblacional (σ)*:* es la raíz cuadrada de la media aritmética de las desviaciones cuadráticas de los valores de la variable, con respecto a la media aritmética de la población (Armas, 1988).

Regla empírica de la distribución normal o distribución de Gauss: lleva este nombre porque está basada en la experiencia y está sujeta a comprobación y comparación de los datos en estudio con los valores de probabilidad teóricos. Los valores obtenidos en la media, la mediana y la moda deben coincidir; la distribución debe ser unimodal (poseer una sola moda). *Está caracterizada por dos parámetros:* la media poblacional (μ) y la desviación típica poblacional (σ), donde la distribución de Gauss (G) se presenta en forma de función de la siguiente manera: $G(\mu, \sigma)$. *Geométricamente,* se interpreta la media poblacional (μ), como un factor de traslación y la desviación típica poblacional (σ), como un factor de escala o grado de dispersión y *probabilísticamente, s*i tomamos intervalos centrados en la media poblacional (μ), y cuyos extremos están una distancia de la desviación típica poblacional (σ), entonces tenemos una probabilidad del 68%; si los extremos están a una distancia de dos desviaciones típicas de la población (2σ), entonces hay una probabilidad del 95% y si los extremos están a distancia de tres desviaciones típicas de la población (3σ), entonces existe una probabilidad del 99% (Levin y otros; 2004).

2.2.4.- *Contexto curricular*

Basados en el currículo elaborado en 1987 el cual se implementó en el año escolar 1987-88 donde se especifican los programas de estudio y manuales del docente del área de Matemática de la Tercera Etapa de Educación Básica y el ciclo de Educación Media Diversificada para los sectores Urbano, Rural, Indígena y Fronteras, a continuación se muestra la unidad en la que se ubica el contenido y objetivo(s) específico(s) de la Estadística incluyendo críticas sobre su estructura:

Séptimo grado:

- ➢ Unidad 6: Estadística.
- Objetivo 30: Distribución de frecuencias.
  Agrupar datos estadísticos en intervalos de clases.
  Determinar la frecuencia absoluta acumulada en una colección de datos agrupados.
- Objetivo 31: Histogramas de frecuencias absolutas.
  Elaborar histogramas de frecuencias absolutas.
  Organizar datos en tablas de distribución de frecuencias.
  Analizar los datos representados en histogramas de frecuencia absoluta.

En este grado, en el objetivo treinta (30) no se menciona que se deben enseñar los conceptos básicos de la Estadística, los cuales son fundamentales para realizar tablas de frecuencia, tampoco se especifica los tópicos que se deben abarcar en la Distribución de Frecuencias, lo cual es elemental para realizar el cuerpo de esta tabla. En el objetivo treinta y uno (31) no aparecen mencionados los gráficos de barras simples y sectores, quitándole la importancia que reviste dicho tópico, debe realizarse mayor hincapié en esto debido a que dichos gráficos son recomendados para variables de tipo cualitativo y cuasi – cuantitativo, además debe ser un conocimiento previo del histograma porque éste es recomendado para variables cuantitativas.

Octavo grado:

- ➢ Unidad 8: Estadística
- Objetivo 27: Media aritmética y moda de datos agrupados.

  Calcular la media aritmética y la moda de una distribución de datos agrupados. Resolver problemas en los que se utilicen la media aritmética y la moda de una distribución de datos agrupados.

En este grado se encuentran bien entrelazados o encadenados los contenidos con los de séptimo grado; además, se presenta bajo una estructura sencilla.

Noveno grado:

- ➢ *Unidad 8: Estadística y Probabilidad*
- Objetivo 26. Estadística y Probabilidad.

  Resolver problemas en los cuales se utilicen nociones elementales de estadística.

Este programa está muy acertado, porque lo que hace es crear en el alumno un aprendizaje significativo, reforzando los contenidos alcanzados durante los años anteriores y utilizando la resolución de problemas como herramienta fundamental.

Segundo año de Educación Media Diversificada

- ➢ *Unidad 5: Métodos numéricos*
- Objetivo 8. Estadística

  Medidas de tendencia central: media, moda y mediana.
  Medidas de dispersión: rangos.
  Cuartiles, deciles y percentiles.
  Varianza de una población.

Varianza de una muestra: desviación.

Regla empírica: curva normal, franja de normalidad.

El programa de segundo año de Educación Media Diversificada, consta de una estructuración mal planteada, ya que no existe una secuencia acorde en la unidad de Estadística referente a los métodos numéricos, esto se ve reflejado en el orden de los objetivos específicos, debido a que es recomendado comenzar con las medidas de posición no central (cuartiles, deciles y percentiles) y luego las medidas de posición central (moda, media y mediana), porque esto es conocimiento previo fundamental para los objetivos posteriores, tales como: el estudio de las medidas de dispersión muéstrales (varianza y desviación típica) y las medidas de dispersión poblacionales y su respectiva regla empírica. Se puede considerar que el ordenamiento planteado es una de las causas de la gran deficiencia que presentan los alumnos en el aprendizaje de la Estadística, contenido importante, ya que de éste depende en gran parte el aprendizaje de otros contenidos posteriores durante sus estudios universitarios. Por ello, se considera necesaria una reordenación de los contenidos programáticos de Estadística para el segundo año de Educación Media Diversificada, incluyendo una manera más pedagógica de enseñar los conceptos básicos de dicha ciencia, y en tal sentido favorecer en el alumno la posibilidad de estudiar y comprender este contenido matemático.

CAPÍTULO 3. Marco metodológico

3.1.- *Tipo de investigación*

De acuerdo a los tipos de investigación que plantea Hurtado (2000) y el grado de profundidad que la investigación presenta, corresponde a una *Investigación Proyectiva*, ya que consiste en el diseño de un Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato que permita superar las fallas detectadas en la revisión biblio-hemerográfica, curricular y diagnóstica para la construcción del aprendizaje de las nociones básicas de la Estadística y dar sentido a la problemática planteada inicialmente en esta investigación.

3.2.- *Diseño de investigación*

Según Sabino (1992) es un diseño *no experimental* porque consiste en desglosar la situación pedagógica para generar información exacta e interpretable, es decir, el diseño de un Módulo de enseñanza – aprendizaje de la Estadística en el bachillerato y es *transversal*, porque se recolectan datos en un momento y tiempo único (Hernández, Fernández y Batista; 2003).

3.3.- *Definición de eventos*

Proceso generador

Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato mediante situaciones problemas.

Evento a modificar

El método de enseñanza-aprendizaje de la Estadística en el bachillerato, con el fin de que los estudiantes comprendan y den solución a situaciones problemas.

3.4.- *Población y muestra*

Población

Está compuesta por el conjunto de profesores de Matemática de bachillerato del Municipio Libertador del Estado Mérida y el conjunto de alumnos de segundo año de Educación Media Diversificada del Municipio Libertador del Estado Mérida.

Muestra

Se tomará al azar una muestra de cinco (5) profesores especialistas en Matemática de bachillerato del Municipio Libertador del Estado Mérida para validar el Módulo y 30 alumnos de segundo año de Educación Media Diversificada del Municipio Libertador del Estado Mérida para realizar la prueba diagnóstica.

3.5.- *Técnicas e instrumentos de recolección de datos*

Para obtener la información necesaria y alcanzar los objetivos de la investigación, se utiliza una prueba diagnóstica de conocimiento (ver Anexo B) con preguntas de selección múltiple y un cuestionario por medio del "Instrumento de validación del Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas" (ver Anexo E), en el mismo se emplea una tabla de doble entrada para evaluar cada clase de acuerdo a seis criterios como son: presentación, secuencia conceptual, ejemplos ilustrados, problemas contextualizados, estrategia didáctica y fundamento matemático siendo cada uno evaluado de la siguiente manera: bueno o regular o malo.

3.6.- *Descripción del procedimiento*

El procedimiento de la presente investigación se describe a través de los siguientes pasos:

<u>Paso 1:</u> Elaborar y validar la prueba diagnóstica de conocimiento.
<u>Paso 2:</u> Aplicar la prueba diagnóstica de conocimiento.
<u>Paso 3</u>*:* Realizar un análisis cuantitativo de los resultados de la prueba diagnóstica de conocimiento.

Paso 4: Diseñar el Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato a través de situaciones problemas.
Paso 5*:* Elaborar el instrumento de validación del Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas.
Paso 6*:* Entregar a los docentes evaluadores el Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas para llevar a cabo el proceso de validación.
Paso 7*:* Realizar un análisis cuali – cuantitativo de los resultados de instrumento de evaluación del Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas.

3.7.- *Tipo de análisis a utilizar*

Se utiliza el Coeficiente de Proporción de Rangos el cual permitirá establecer la validez de contenido de cada ítem, la validez de contenido de todo el instrumento y el nivel de concordancia entre los jueces el cuál arrojó lo siguiente $CPRtc = CPRt - Pe = 0{,}978 - 0{,}036 = 0{,}942$ (ver Anexo D). Además se emplean técnicas de análisis descriptivas, efectuando un análisis cuantitativo (estadístico) de manera comparativa y explicativa a los resultados de datos recabados por la prueba diagnóstica de conocimiento. La validación del Módulo a través del instrumento elaborado, fue realizada por cinco (5) docentes de Matemática especialistas en el área, lo cuales consideraron que el 100% del mismo estuvo bueno en cuanto a los seis criterios evaluativos.

CAPÍTULO 4. Resultados del diagnóstico

4.1.- *Antecedentes del estudio*

Se efectuó un análisis detallado en tres (3) instituciones diferentes del Municipio Libertador de la ciudad de Mérida, revisando los planes de lapso en el área de Matemática de los últimos tres (3) años, lo cual permitió detectar que solo en un (1) lapso de los mismos fue incluido la Estadística, por tanto la realidad es que el contenido objeto de estudio, no está siendo incluido en los planes de lapso. Debido a esto se realizó una entrevista estructurada a tres (3) profesores especialistas con más de 10 años de experiencia en el área de Matemática del Municipio Libertador de la ciudad de Mérida la cual arrojó lo siguiente:

En la primera pregunta sobre si habían enseñado Estadística durante el tiempo que tenían dando clases, dos de los profesores entrevistados respondieron que no porque consideran más importantes otros contenidos del programa, además no enseñan esta ciencia porque está al final del currículo de estudio y no les alcanza el tiempo del año escolar. El otro profesor respondió que sólo lo enseñó en el segundo año de Educación Media Diversificada durante el año escolar 2007-2008. La siguiente pregunta trató sobre las dificultades más comunes que encontraron en la enseñanza de la Estadística, sólo un docente argumentó que los alumnos no tienen los conocimientos previos de la Estadística necesarios para la comprensión de un contenido posterior y de mayor dificultad, los otros dos no respondieron dado que nunca habían enseñado Estadística.

A pesar de enseñar o no estadística, se les preguntó si consideraban importante la enseñanza de los contenidos de Estadística en el bachillerato, todos los entrevistados manifestaron que la enseñanza de la Estadística en el bachillerato es muy importante y por lo tanto se debería enseñar, ya que contribuyen en el desenvolvimiento de la realidad circundante, para el desarrollo de proyectos de estudios exigidos por otras asignaturas tales como Biología, Física y otras, y porque se utilizarán posteriormente en estudios universitarios. Sin embargo, explicaron que no lo enseñan porque no existen alternativas innovadoras e interesantes para su enseñanza, por ello dejan de lado la Estadística.

Por último se les solicitó la opinión acerca de cómo consideran las estrategias de enseñanza de la Estadística sugerida en los libros de textos, a lo que respondieron los tres docentes que los libros de textos están un poco desactualizados, por tanto solo se deberían tomar como una orientación del contenido a impartir y los ejemplos que están allí nada tienen que ver con el entorno, es decir, están totalmente descontextualizados de la realidad.

4.2.- *Diagnóstico de necesidades*

Los docentes entrevistados no enseñan la Estadística debido a que ellos consideran más significativos otros contenidos del programa y esto se ve reflejado en los planes de lapso donde el mismo no está siendo incluido. Tampoco se enseña esta ciencia porque está ubicada al final del currículo de estudio y no les alcanza el tiempo del año escolar por ello sólo en un año se incluyó su enseñanza. El docente que enseñó Estadística en este año escolar, manifestó que los alumnos no tienen los conocimientos previos necesarios para la comprensión de un contenido posterior y de mayor dificultad, lo cual hace observar la falta de aprendizaje de los estudiantes corroborado por el hecho de que la Estadística no se incluye en los planes de lapso. Además todos los entrevistados manifestaron que la enseñanza de la Estadística en el bachillerato es primordial y por lo tanto se debería enseñar, ya que contribuyen en el desenvolvimiento de la realidad circundante, para el desarrollo de proyectos de estudios exigidos por otras asignaturas y porque se utilizarán posteriormente en estudios universitarios. Sin embargo, explicaron que no lo enseñan porque no existen alternativas innovadoras e interesantes en los libros de textos por encontrarse descontextualizados, por ello dejan de lado la Estadística en los planes de lapso.

4.3.- *Evaluación de las condiciones actuales y explicaciones tentativas*

Los docentes de Matemática no imparten los contenidos de la Estadística debido a que consideran más significativos otros contenidos del programa lo cual se ve reflejado en los planes de lapso donde no incluyen estos contenidos. Tampoco enseñan esta ciencia debido a que sus contenidos están al final del currículo de estudio y no les alcanza el tiempo del año escolar, los libros están descontextualizados, también podría deberse a que los docentes no tienen el conocimiento sobre la misma.

4.4.- *Posibles tendencias futuras*

Los estudiantes no tendrán los conocimientos sobre esta ciencia porque los docentes no la enseñan, esto trae como consecuencia que los estudiantes no aprenden los contenidos de esta ciencia. Por lo tanto, no se logrará capacitar a estudiantes para un curso básico de Estadística a nivel universitario, además no lograrán interpretar y evaluar críticamente la información Estadística que se puede encontrar en diversos contextos, incluyendo los medios de comunicación, y la vida diaria, tampoco podrían discutir o comunicar sus opiniones respecto a tales informaciones Estadísticas cuando sea relevante.

4.5.- *Síntesis diagnóstica*

La información recabada resalta la gran deficiencia que existe en la enseñanza – aprendizaje de la Estadística en el bachillerato debido a que los docentes especialistas en el área no imparten el contenido porque no le dan el peso necesario y por ubicarse al final del currículo escolar, no lo enseñan porque le falta tiempo por todo lo antes dicho no lo están incluyendo en los planes de lapso, por lo tanto los estudiantes no se preparan y no aprenden lo básico de la Estadística.

CAPÍTULO 5. Presentación de la propuesta

5.1.- *Justificación de la propuesta.*

El interés por realizar un Módulo para la enseñanza de la Estadística radica en los resultados que arrojó la aplicación de una prueba diagnóstica para determinar los conocimientos previos de los estudiantes del Segundo año de Educación Media Diversificada, ya que éstos debieron aprender una serie de nociones básicas de Estadística. La prueba se aplicó en cinco (5) instituciones diferentes del Municipio Libertador de la ciudad de Mérida, se tomaron al azar treinta (30) alumnos, seis (6) en cada institución, la misma arrojó índices alarmantes tales como: veintitrés (23) alumnos reprobaron lo cual representa el 76,7% y el promedio de notas fue de 7 puntos, no tienen claro los conceptos básicos de Estadística (Estadística, población, muestra, dato) porque el 80% no respondió correctamente, no diferencian los tipos de variable ya que 28 de los 30 evaluados respondió incorrectamente, el 76% de los estudiantes no diferencian los tipos de frecuencia y el 83% no lograron calcular e interpretar las medidas de tendencia central.

Los anteriores índices alarmantes hacen notar la falta de preparación y aprendizaje de los contenidos básicos de la Estadística en los estudiantes del bachillerato, además de no comprender situaciones estadísticas de su entorno; por ello, se plantea el Módulo a través de situaciones problemas contextualizadas, que permita formar futuros ciudadanos con capacidad lectora e interpretación de tablas y gráficos estadísticos que con frecuencia aparecen en los medios informativos. También es útil para estudios ulteriores, ya que en muchas profesiones se precisan conocimientos básicos del tema y su estudio ayuda al desarrollo personal, fomentando un razonamiento crítico basado en la valoración de la evidencia objetiva.

5.2- *Finalidad y metas de la propuesta.*

Finalidad

Facilitar a los docentes de Matemática un Módulo como una alternativa diferente e innovadora de las nociones básicas de la Estadística en la Tercera Etapa de Educación Básica y el segundo año del ciclo de Educación Media Diversificada que permita a los estudiantes la construcción de su aprendizaje logrando calcular, comprender e interpretar activamente esta ciencia.

Metas.

- Utilizar el Módulo como material de apoyo para la preparación de las clases de la Estadística de acuerdo al contenido a enseñar.
- Incorporar a los estudiantes en la enseñanza – aprendizaje de las nociones básicas de la Estadística.
- Lograr que el estudiante relacione o vincule la Estadística con la realidad, al exponerlos a distintas ideas, perspectivas y situaciones problemas, que le ayuden a construir, calcular, comprender e interpretar con claridad.

5.3.- Presentación de la propuesta

5.3.1.- *Descripción de la propuesta*

Tabla 1. Esquema general de la Propuesta de Orientación Didáctica

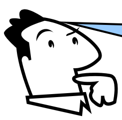
<table>
<tr><th>Clases</th><th>Contenido</th><th colspan="3">Tópicos</th><th>Nº de Página</th></tr>
<tr><td rowspan="8">Clase 1</td><td rowspan="8">Conceptos básicos.</td><td colspan="3">Estadística</td><td>45</td></tr>
<tr><td colspan="3">Población</td><td>46</td></tr>
<tr><td colspan="3">Muestra</td><td>46</td></tr>
<tr><td colspan="3">Dato</td><td>47</td></tr>
<tr><td rowspan="4">Variables estadísticas</td><td colspan="2">Variables cualitativas</td><td>48</td></tr>
<tr><td colspan="2">Variables cuasi-cuantitativas</td><td>49</td></tr>
<tr><td rowspan="2">Variables cuantitativas</td><td>Discretas</td><td>49</td></tr>
<tr><td>Continuas</td><td>50</td></tr>
<tr><td rowspan="5">Clase 2</td><td rowspan="5">Organización de datos para variables cualitativas.</td><td colspan="3">Categorías</td><td>53</td></tr>
<tr><td colspan="3">Tabla</td><td>53</td></tr>
<tr><td rowspan="3" colspan="2">Frecuencia</td><td>Frecuencia absoluta</td><td>53</td></tr>
<tr><td>Frecuencia relativa simple</td><td>54</td></tr>
<tr><td>Porcentaje</td><td>55</td></tr>
<tr><td rowspan="3">Clase 3</td><td rowspan="3">Organización de datos para variables cuasi cuantitativas.</td><td colspan="3">Frecuencia acumulada</td><td>59</td></tr>
<tr><td colspan="3">Frecuencia relativa acumulada</td><td>60</td></tr>
<tr><td colspan="3">Porcentaje acumulado</td><td>60</td></tr>
</table>

Clase 4	Organización de datos para variables cuantitativas o distribución de frecuencias.	Rango o recorrido	64
		Número de clases	64
		Amplitud de la clase o intervalo	64
		Límite aparente inferior	65
		Limite aparente superior	65
		Distribución de frecuencias	65
Clase 5	Representación gráfica de las variables cualitativas y cuasi – cuantitativas.	Gráfico de sectores o circular	68
		Gráfico de barras simples	70
Clase 6	Representación gráfica de variables cuantitativas.	Histograma	73
		Límite real inferior	74
		Límite real superior	74
		Marca de clase	74
Clase 7	Medidas de tendencia no central.	Medidas de tendencia no central	78
		Percentiles	79
		Deciles	81
		Cuartiles	84
Clase 8	Medidas de tendencia central.	Medidas de tendencia central	87
		Moda	87
		Mediana	90
		Media	92
Clase 9	Medidas de dispersión absolutas.	Rango	94
		Rango percentílico	95
		Rango cuartílico	96
Clase 10	Medidas de dispersión absolutas alrededor de la media.	Varianza muestral	98
		Varianza poblacional	98
		Desviación típica muestral	99
		Desviación típica Poblacional	99
		Regla empírica de la distribución normal	100
Clase 11	Medidas de dispersión absolutas basadas en un valor central determinado.	Desviación media	104
		Desviación mediana	105
Clase 12	Excel.	Cálculo en Excel	108

Propuesta de Orientación Didáctica para la enseñanza-aprendizaje de la Estadística en el bachillerato mediante la solución de situaciones problemas.

Clase 1: Conceptos básicos.

En la vida diaria se pueden contabilizar muchas situaciones o problemas propios de nuestra realidad; en el caso de ser estudiantes se pueden contar cuántas damas y caballeros hay en un salón de clases, cuál es la edad de cada uno, u otra situación problemática que se pudiera describir en la sociedad como por ejemplo: el embarazo precoz, las drogas, la delincuencia, la contaminación ambiental o cualquier otro problema que se nos pueda ocurrir.



El objetivo de la clase es buscar que el alumno utilice problemas de la vida diaria en el aprendizaje de conceptos básicos de la Estadística.

Problema 1: Los estudiantes
Un profesor investigador del Colegio Salesiano "San Luis" (Mérida - Venezuela) durante el presente año, quiere indagar si la edad (en años cumplidos) , la clase social (alta, media, baja), el género (femenino, masculino), el nivel educativo de la madre (primaria, secundaria, superior); tienen influencia en el promedio aritmético de las notas en puntos de los estudiantes del Colegio Salesiano "San Luis"; para ello toma al azar una sección de segundo año del ciclo de Educación Media Diversificada, la cual esta conformada por 32 estudiantes, dichos datos los recopila entre la Coordinación Docente y una encuesta en la institución.

Lo que hace el profesor en el problema 1 es:

- Tomar los estudiantes según el género, clase social, nivel educativo de la madre y promedio aritmético de notas; es decir, está clasificando la información.
- Contar cuántos estudiantes hay del género masculino y cuántos del género femenino; es decir, está recogiendo información, también la está resumiendo y hallando regularidades entre la edad, la clase social, el género y el promedio de notas.

Lo que está haciendo el profesor es lo que hace la Estadística, que es una ciencia que a través de métodos y procedimientos recoge, clasifica, resume, halla regularidades y analiza la información.

- Al determinar si tiene influencia la edad, la clase social y el género en el promedio aritmético, se está analizando la información.

A partir de lo anterior, se puede establecer que: al realizar un conteo del número de estudiantes que tienen 16 años o de alumnos de clase media, se está realizando una forma sencilla de Estadística.

Con esto, lo que se quiere es enfocar el paso evolutivo de esta ciencia

Esto que está haciendo el profesor del Colegio Salesiano "San Luis" siempre lo ha hecho el ser humano, aunque no recibía el nombre de Estadística, ya desde los comienzos de las civilizaciones han existido formas sencillas de Estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas.

El objetivo es inducir conceptos básicos a través de la historia.

Se puede ver que hasta en la Biblia está inmerso el concepto de la Estadística. Los libros bíblicos de números y crónicas incluyen, en algunas partes, trabajos de Estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. Sin embargo sólo muy recientemente la Estadística ha adquirido la categoría de ciencia.

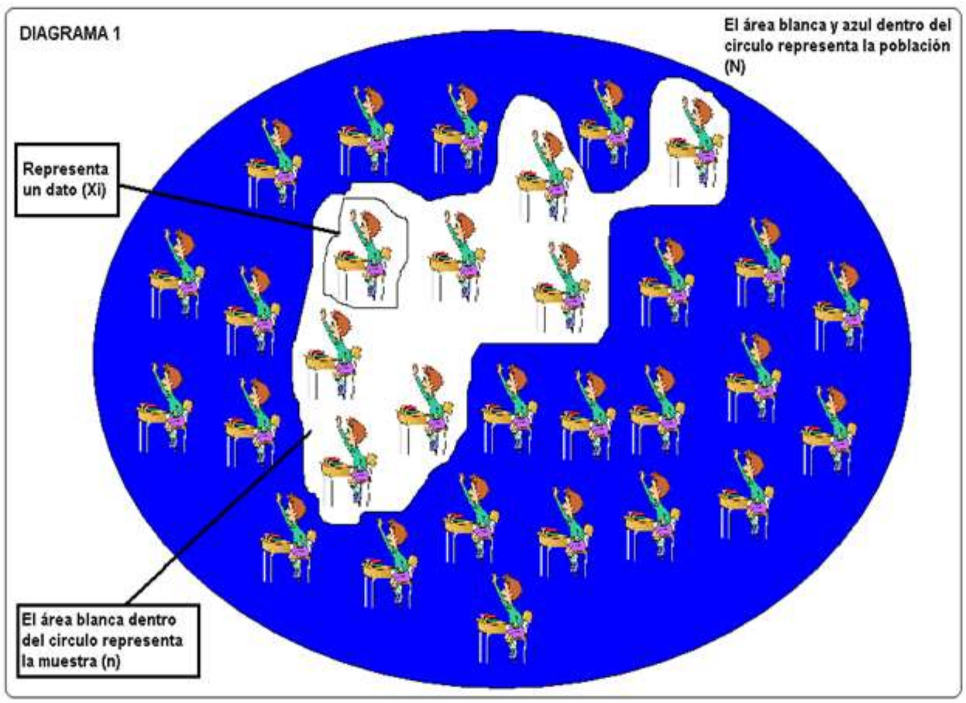
En el problema 1, el profesor está haciendo una indagación a los estudiantes del Colegio Salesiano "San Luis" del estado Mérida, los cuales tienen en común, que son estudiantes de la misma institución, a esto en términos estadísticos se le llama población. Es decir, la población es a quién le vamos a indagar sobre un tema con propiedades comunes; de manera formal, **población (N):** es el conjunto de elementos que cumplen ciertas propiedades comunes.

En el problema 1 cuando el profesor toma al azar los estudiantes de una sección del segundo año del ciclo diversificado, está explorando un fragmento o pedazo de la población, y a esto le llamamos **muestra (n)**, es decir, una muestra (o parte) de la población es un subconjunto de elementos que cumplen ciertas propiedades comunes.

En el problema 1, cuando el profesor busca información de cada uno de los 32 estudiantes del Colegio Salesiano "San Luis" del estado Mérida, está tomando en cuenta a cada uno de los **datos** (x_i), que son cada uno de los individuos, cosas, entes abstractos que integran una población determinada.

Con este diagrama se engloban los conceptos de población, muestra y dato.

Veamos el Diagrama 1:



Problemas de consolidación

1. De los siguientes enunciados identifique: población, muestra y dato.

- Con la finalidad de estimar las preferencias electorales en una elección nacional, se realiza un sondeo de opinión en un estado con 130.000 electores. Para este fin, se entrevista a 1300 de los electores, mediante una selección aleatoria.

- Se realizó un estudio para conocer la opinión de la comunidad rural de Bailadores que posee 3.500 habitantes, respecto a sí aceptar o no un programa de alfabetización de adultos. Para llevar a cabo este estudio, se consultó a 500 habitantes de Bailadores.

- En la Urbanización “Carlos Sánchez” (Ejido – Mérida), se está realizando una encuesta con el propósito de lograr obtener un nuevo ambulatorio para la comunidad, para realizar esta investigación se entrevistó los habitantes de las calles 1, 2, 3, 4 y 5 de dicha urbanización.

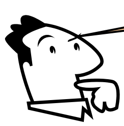
Lo curioso de la Estadística

Un reciente estudio psicopedagógico ha mostrado que los niños de pie grande saben leer mejor que los de pie pequeño. ¿Permitirá el tamaño del pie medir la capacidad de lectura de los niños?

Con esta actividad se explica, que la Estadística no está aislada del mundo real, es decir, los problemas son susceptibles a ser medidos.

No, desde luego. El estudio se hizo sobre escolares que están en crecimiento. Todo cuanto se demostró en el estudio es que los niños mayores, cuyos pies son más grandes, leen mejor que los niños menores.

Sigamos desarrollando los conceptos básicos de la Estadística

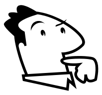


El objetivo es apelar a los conocimientos previos para definir conceptos importantes para el proceso de aprendizaje.

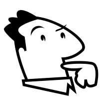
En el problema 1 la edad, la clase social, el género, el nivel de instrucción de la madre y el promedio de notas son propiedades, rasgos o cualidades de los elementos de la población, que se van a estudiar a través de la muestra y que se llaman **variables estadísticas**; es decir, que poseen variedad.

Observen que la información aportada por el problema 1, nos ayudará a definir los diferentes tipos de variables:

La variable género, se divide en masculino y femenino, y cada alumno tiene bien claro su género, a éste tipo de variables se le llaman **variables cualitativas:** son aquellas variables que como su nombre lo indica están relacionadas con las características exteriores propias de cada uno de los individuos en estudio.

 **EJEMPLOS:** estado civil, número de cedula de identidad, color de ojos, opinión sobre el programa ¿Quién quiere ser millonario? (bueno o malo), marcas de computadoras, ¿Posee empleo? (si o no), e- mail de cada persona, etc.

En el problema 1 existe una variable llamada clase social y otra llamada nivel educativo de la madre, ambas están ordenadas (ver problema1), cuando llevan un orden, se denominan **variables cuasi-cuantitativas**: son variables que se pueden ordenar entre las categorías de la variable.

 **EJEMPLOS:** rangos militares (están ordenados desde el soldado hasta general), opinión sobre el programa ¿quién quiere ser millonario? (bueno, regular, malo), los meses del año, opinión sobre el servicio prestado por el servicio de cantina (excelente, muy bueno, regular, malo, muy malo).

Teniendo en cuenta los conocimientos previos, la variable edad está expresada en unidades de tiempo (segundos, minutos, días, meses, años) y la variable promedio aritmético de notas en puntos, además de ser cuantificables, a estas variables se le denomina **variables cuantitativas**: son caracterizadas por alguna información numérica que se le puede asociar a los individuos de una población o muestra**.**

Todas las edades en meses cumplidos vienen expresadas en valores enteros, tal como la vamos a presentar en la variable edad de los estudiantes del problema 1, éstos valores son llamados en Matemática **discretos**, y cada variable cuantitativa que presenta los datos de esa manera en Estadística lleva por nombre **variable cuantitativa discreta**, a partir de esto, definimos este tipo de variables como aquellas, cuyos datos están expresados por números enteros, es decir, entre dos números consecutivos no hay valores intermedios.

En el problema 1 a los estudiantes se les investiga la variable nota promedio, de acuerdo a los conocimientos previos, se sabe que un estudiante puede tener una nota promedio de 13,2 puntos, y cada variable cuantitativa que muestra los datos de este modo, en

Estadística lleva por nombre, **variable cuantitativa continua**, a partir de esto, definimos este tipo de variables como aquellas, cuyos datos están expresados por números reales, lo cual representa que entre dos valores consecutivos existen valores intermedios.

Problemas de consolidación

1. Con la información aportada por el diagrama 2 identifique los tipos de variables.

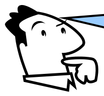
DIAGRAMA 2

2. Marca con una X la opción correspondiente al tipo de variable:

Variables	Cualitativa	Cuasi - cuantitativa	Cuantitativa Continua	Cuantitativa Discreta
Días de la semana.				
Nacionalidad de los residentes de Venezuela				
Número de la camiseta de los jugadores de baloncesto.				
Cantidad de dinero que gasta diariamente en desayunar un estudiante del Colegio Salesiano “San Luis”				
Color exterior que presenta una fruta cuando está madura.				
Calificación obtenida por un estudiante al concluir el noveno grado de Educación Básica.				
El nivel de satisfacción con un trabajo (medido en escala Likert: muy bueno, bueno, regular, malo, muy malo).				

Estado civil de una persona que contrae nupcias por segunda vez.				
Tiempo empleado por un estudiante en llegar desde su casa hasta el liceo.				
Número de días de los meses del año.				
La cantidad de agua consumida por una familia semanalmente.				
Opinión sobre el programa quién quiere ser millonario.				
Edad (años).				
La cantidad de veces que una persona fue al cine en el último año.				
La puntuación que obtuvo un estudiante en una prueba de Matemática.				
La concentración en bachillerato (medido en escala Likert: muy aceptable, aceptable, poco aceptable, nada aceptable).				

<u>Clase 2</u>: Organización de datos para variables cualitativas.



El objetivo de la clase es buscar que el alumno utilice variables cualitativas y organice los datos mediante tablas.

Los datos que se presentan a continuación en la **TABLA A** son de los 32 estudiantes del segundo año del ciclo de Educación Media Diversificada tomados al azar del Colegio Salesiano "<u>San Luis</u>" del estado Mérida, según las variables **edad (en meses cumplidos), clase social (alta, media, baja), género (masculino, femenino), nivel de instrucción de la madre (primaria, secundaria, superior),** para encontrar la relación de las variables anteriores con el **promedio aritmético de notas (puntos).**

TABLA A

Nº	EDAD (MESES CUMPLIDOS)	CLASE SOCIAL	GÉNERO	NIVEL DE INSTRUCCIÓN (MADRE)	PROMEDIO DE NOTAS (PUNTOS)
1	205	Alta	Masculino	Superior	16,56
2	196	Media	Femenino	Superior	18,21
3	195	Media	Femenino	Primaria	13,44
4	198	Baja	Femenino	Secundaria	12
5	180	Media	Masculino	Secundaria	14,1
6	206	Baja	Femenino	Primaria	13,87
7	216	Alta	Masculino	Secundaria	10,35
8	218	Baja	Masculino	Secundaria	11,2
9	209	Baja	Femenino	Primaria	12,72
10	194	Baja	Femenino	Secundaria	17,9
11	182	Media	Masculino	Superior	16
12	192	Baja	Femenino	Primaria	12,59
13	200	Alta	Masculino	Superior	18,5
14	195	Baja	Femenino	Primaria	19,49
15	194	Media	Masculino	Superior	13,88
16	181	Baja	Masculino	Superior	15,39
17	199	Media	Femenino	Secundaria	16,23
18	198	Baja	Femenino	Primaria	14,73
19	208	Baja	Femenino	Primaria	14,7
20	188	Baja	Masculino	Secundaria	10
21	212	Baja	Femenino	Primaria	18
22	184	Baja	Femenino	Secundaria	16,9
23	195	Media	Femenino	Superior	15,64
24	193	Baja	Femenino	Secundaria	14,85
25	193	Alta	Masculino	Superior	16,3
26	196	Media	Femenino	Superior	14,1
27	199	Baja	Masculino	Secundaria	15,72
28	198	Baja	Femenino	Secundaria	16,68

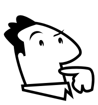
29	207	Media	Masculino	Secundaria	14,1
30	204	Alta	Femenino	Secundaria	15,79
31	201	Baja	Femenino	Primaria	11,45
32	196	Baja	Masculino	Secundaria	12,33

Se puede ver que la ordenación de datos presentada en la **TABLA A** es sumamente extensa, por ello el profesor del problema 1, los clasificará de acuerdo a sus categorías, que es cada una de las clases establecidas en una profesión, carrera o actividad y en nuestro problema, por ejemplo la variable género se agrupa en femenino y masculino, la variable clase social en alta, media y baja, y la variable nivel de instrucción de la madre, en primaria, secundaria y superior, a partir de lo anterior se define **categorías**: como cada uno de los grupos básicos en los que puede incluirse o clasificarse cada variable, bien sea cualitativa o cuasi-cuantitativa.

Entonces vamos a resumir los datos anteriores por medio de una **tabla,** consiste en presentar los datos organizadamente en arreglos tabulares. La presentación tabular constituye la forma más general de presentar datos estadísticos, para esto tomemos la variable género (variable cualitativa):

Para ello se cuentan los estudiantes, que hay del género masculino y del género femenino, y tenemos que existen 13 masculinos y 19 femeninos, cuando hacemos este conteo, estamos determinando la frecuencia o números de elementos con que se repite cada categoría, a esto se le llama **frecuencia absoluta** (n_i): es el número de veces que se repite un dato (x_i) y se puede hallar para cualquier tipo de variable.

Se debe hacer referencia que sirve para variables cualitativas, cuasi-cuantitativas y cuantitativas continuas y discretas.



Veamos cómo quedará la tabla con lo que hemos aprendido:

Género	n_i
Masculino	13
Femenino	19
Total	32

Observen que identificamos en la primera fila el nombre de la variable y la frecuencia absoluta, luego se anotan las categorías y por último el total correspondiente a la suma de n_i, éste debe ser igual al total de la muestra.

Interpretación:

De acuerdo a la información de la tabla se puede concluir, que existen más estudiantes del género femenino que estudiantes del género masculino.

Se divide la cantidad de datos por categorías entre el total de la muestra, y se tiene:

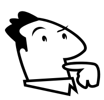
Masculino: 13/32=0,41.

Femenino: 19/38=0,59.

Siempre se toman dos cifras decimales

Interpretación: de acuerdo a los cálculos efectuados se tiene que trece de treinta y dos o 0,41 de los estudiantes son de género masculino y diecinueve de treinta y dos o 0,59 son del género femenino.

En la ciencia en estudio lleva por nombre**,** la **frecuencia relativa simple** (f_i): es una proporción del número de datos que se repiten entre el total de datos de la muestra; es decir, es el cociente de la frecuencia absoluta (n_i) entre el total de datos de la muestra (n), se puede hallar para cualquier tipo de variable y se calcula así: $f_i = \frac{n_i}{n}$.



Veamos como quedará la tabla con lo que hemos aprendido:

Género	n_i	f_i
Femenino	19	0,59
Masculino	13	0,41
Total	32	1

Observen que identificamos en la primera fila el nombre de la variable y las frecuencias calculadas (n_i, f_i) luego se anota las categorías y por último el total correspondiente a la suma de las frecuencias.

Nota: la suma de las frecuencias relativas absolutas siempre debe ser igual a uno.

A continuación se pueden multiplicar el resultado de la frecuencia relativa simple por cien, y esto lleva por nombre porcentaje, y se calcula así:

Masculino: (13/32) x 100= 41 %.

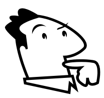
Femenino: (19/32) x 100= 59%.

El porcentaje siempre lo vamos a escribir con el símbolo %.

Interpretación: se puede observar que el cincuenta y nueve por ciento o 59% de los estudiantes son del género femenino, lo que representa 59 de 100 partes totales y el cuarenta y uno por ciento o 41% de los estudiantes son del género masculino, lo que representa 41 de 100 partes totales.

En consecuencia, de lo anterior se puede definir **porcentaje (%)** como una forma de expresar una proporción o fracción con un cociente de denominador cien; es decir, el producto de la frecuencia relativa simple (f_i) multiplicada por cien, se puede hallar para cualquier tipo de variable y se calcula así:

$$\% = \frac{n_i}{n} \times 100 = f_i \times 100.$$



Veamos como quedará la tabla usando el porcentaje:

Género	ni	fi	%
Masculino	16	0,42	42
Femenino	22	0,58	58
Total	38	1	100

Observen que identificamos en la primera fila el nombre de la variable, las frecuencias y porcentajes calculados, luego se anota las categorías y por último el total correspondiente a la suma de las frecuencias y porcentajes.

Nota: la suma de los porcentajes siempre debe ser igual a cien.

Para completar la tabla anterior se deben agregar los siguientes elementos:

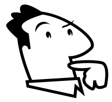
El objetivo de esta analogía es enfocar la importancia de que la tabla tenga todos sus elementos.

- **Una tabla sin título, es como un ser humano sin cabeza.**
- **Una tabla sin cuerpo es como un ser humano sin tronco.**
- **Una tabla sin fuente es como un ser humano sin piernas.**

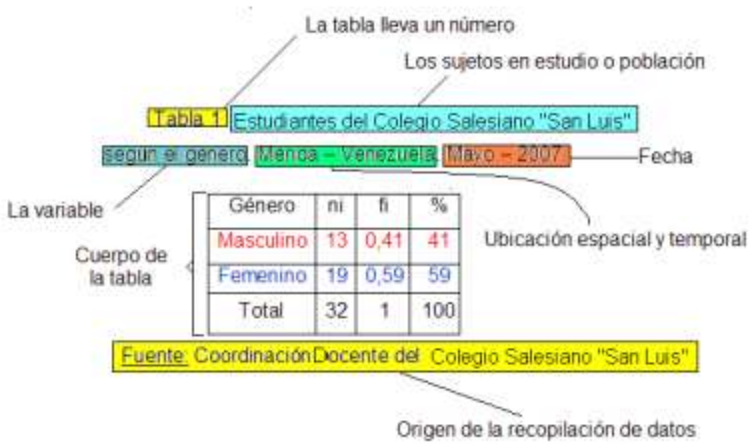
Tabla 1. Estudiantes del Colegio Salesiano "San Luis" según el género. Mérida – Venezuela. Mayo – 2007.

Género	ni	Fi	%
Masculino	13	0,41	41
Femenino	19	0,59	59
Total	32	1	100

Fuente: Coordinación Docente del Colegio Salesiano "San Luis".



Veamos esta tabla con más detenimiento:



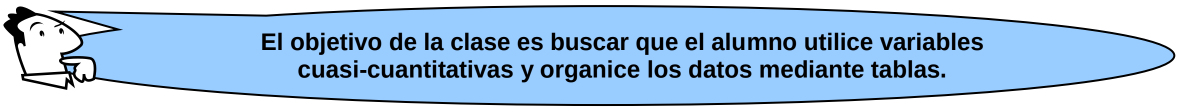
Problemas de consolidación

Suponga que los siguientes datos fueron recolectados a estudiantes de séptimo grado del Colegio Salesiano “san Luis”, Mérida – Venezuela durante el mes de mayo de 2007.

1. Construya una tabla con todos sus elementos para cada una de las variables e interprete los resultados:

Nº	Color de ojos	Opinión sobre el programa ¿Quién quiere ser millonario?
1	Negro	Bueno
2	Marrón	Bueno
3	Gris	Malo
4	Negro	Bueno
5	Negro	Malo
6	Marrón	Bueno
7	Negro	Bueno
8	Gris	Bueno
9	Negro	Malo
10	Gris	Malo
11	Marrón	Bueno
12	Marrón	Malo
13	Marrón	Bueno
14	Negro	Malo
15	Negro	Bueno
16	Negro	Malo
17	Gris	Bueno
18	Negro	Malo
19	Marrón	Bueno
20	Marrón	Malo

<u>Clase 3</u>: Organización de datos para variables cuasi-cuantitativas.



Los datos que se presentan a continuación son de los 32 estudiantes del segundo año del ciclo de Educación Media Diversificad tomados al azar del Colegio Salesiano "<u>San Luis</u>" del estado Mérida mostrado en la **TABLA A** de la clase 2, según las variables **clase social (alta, media, baja)** y **nivel de instrucción de la madre (primaria, secundaria, superior).**

Nº	CLASE SOCIAL	NIVEL DE INSTRUCCIÓN (MADRE)
1	Alta	Superior
2	Media	Superior
3	Media	Primaria
4	Baja	Secundaria
5	Media	Secundaria
6	Baja	Primaria
7	Alta	Secundaria
8	Baja	Secundaria
9	Baja	Primaria
10	Baja	Secundaria
11	Media	Superior
12	Baja	Primaria
13	Alta	Superior
14	Baja	Primaria
15	Media	Superior
16	Baja	Superior
17	Media	Secundaria
18	Baja	Primaria
19	Baja	Primaria
20	Baja	Secundaria
21	Baja	Primaria
22	Baja	Secundaria
23	Media	Superior
24	Baja	Secundaria
25	Alta	Superior
26	Media	Superior
27	Baja	Secundaria
28	Baja	Secundaria
29	Media	Secundaria
30	Alta	Secundaria
31	Baja	Primaria
32	Baja	Secundaria

Ahora se resumen los datos anteriores por medio de una tabla para la variable clase social que por ser una variable cuasi-cuantitativa se le pueden calcular otro tipo de frecuencias y porcentajes, se calcula primero la frecuencia absoluta, frecuencia relativa absoluta y el porcentaje para repasar la clase anterior.

Comencemos con la frecuencia absoluta (n_i), recordemos que es contar cuántos datos existen por cada categoría, en el problema 1, hay 5 estudiantes de la clase social alta, 9 estudiantes de clase social media y 18 de clase social baja.

Sigamos con la frecuencia relativa simple (f_i), recordemos que $f_i = \frac{n_i}{n}$ entonces tenemos que ésta frecuencia para cada una de las categorías quedará del siguiente modo:

Alta: (5/32) = 0,16.

Media: (9/32) = 0,28.

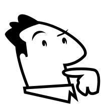
Baja: (18/32) = 0,56.

Lo último que habíamos calculado fue el porcentaje (%), recordemos que $\% = \frac{n_i}{n} \times 100 = f_i \times 100$, entonces el porcentaje para cada una de las categorías quedará del siguiente modo:

Alta: (5/32) x 100 = 16%.

Media: (9/32) x 100 = 28%.

Baja: (18/32) x 100 = 56%.



Veamos cómo quedará la tabla con lo que hemos aprendido:

Clase social	Ni	fi	%
Alta	5	0,16	16
Media	9	0,28	28
Baja	18	0,56	56
Total	32	1	100

Es importante que se encuentren ordenadas las categorías de las variables cuasi-cuantitativas bien sea en orden descendente como están en la tabla o en orden ascendente.

Se puede ver en la tabla anterior, que las categorías se encuentran ordenadas y por esto podemos, sumar o acumular la primera frecuencia absoluta (n_i) con la siguiente y así sucesivamente hasta la última categoría, que debe ser igual al total de la muestra(n). Tomando en cuenta lo anterior, podemos interpretar que hay **5** estudiantes de clase social alta, **14** de clase social alta o media y **32** o el total de la muestra de clase social alta o media o baja.

<u>Es decir:</u>

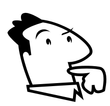
$N_1 = n_{\text{alta}} = n_1 =$ **5.**

$N_2 = n_{\text{alta o}}\ n_{\text{media}} = n_1 + n_2 = 5 + 9 =$ **14.**

$N_3 = n_{\text{alta o}}\ n_{\text{media o}}\ n_{\text{baja}} = n_1 + n_2 + n_3 = 5 + 9 + 18 =$ **32.**

Esta acumulación de frecuencias absolutas lleva por nombre **<u>frecuencia acumulada</u>**(N_i)**:** es la frecuencia de cada dato (x_i) más la suma de los valores anteriores a dicha acumulación, estas frecuencias son válidas solo para variables cuasi–cuantitativas y cuantitativas, se calcula así:

$$N_i = n_1 + n_2 + \ldots + n_i = \sum_{j=1}^{i} n_j .$$



<u>Se representa de la siguiente manera:</u>

Clase social	ni	N_i
Alta	5	5
Media	9	14
Baja	18	32
Total	32	No Lleva

Esta acumulación de datos también se puede realizar con la frecuencia relativa absoluta (f_i), se hace sumando la primera frecuencia relativa simple (f_i) con la siguiente y así sucesivamente hasta la última categoría, que debe ser igual a uno. Tomando en cuenta lo anterior, interpretamos éste tipo de frecuencias así: la proporción de estudiantes de clase social alta es de **0,16**, la proporción de estudiantes de clase social alta o media es de **0,44** y la proporción de estudiantes de clase social alta o media o baja es de **1**; por tanto, tenemos lo siguiente:

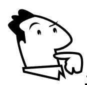
$F_1 = f_1 = f_{\text{alta}} = f_1 =$ **0,16.**

$F_2 = f_{\text{alta o}}\ f_{\text{media}} = f_1 + f_2 = 0{,}16 + 0{,}28 =$ **0,44.**

$F_3 = f_{\text{alta o}}\ f_{\text{media o}}\ f_{\text{baja}} = f_1 + f_2 + f_3 = 0{,}16 + 0{,}28 + 0{,}56 =$ **1.**

Esta acumulación de frecuencias relativas absolutas lleva por nombre **frecuencia relativa acumulada (F_i):** es una proporción del número de datos acumulados entre el total de datos de la muestra; es decir, el cociente de la frecuencia acumulada entre el total de la muestra, solo para variables cuasi–cuantitativas y cuantitativas, se calcula así:

$$F_i = \frac{N_i}{n} = \frac{n_1 + n_2 + \ldots + n_i}{n} = f_1 + f_2 + \ldots + f_i = \sum_{j=1}^{i} f_i .$$

Se representa de la siguiente manera:

Clase social	f_i	F_i
Alta	0,16	0,16
Media	0,28	0,44
Baja	0,56	1
Total	1	No Lleva

Además, se pueden acumular los porcentajes, sumando el primero con el siguiente y seguir así hasta el último, el cual debe dar como resultado 100%. A partir de lo mencionado anteriormente, se interpreta éste porcentaje así: el porcentaje de estudiantes de clase alta es de **16%**, el porcentaje de estudiantes de clase alta o media es de **44%**, el porcentaje de estudiantes de clase alta o media o baja es de **100%** y se puede expresar de la siguiente manera:

P.A.$_1$ = % $_{\text{alta}}$ = %$_1$ = **16%.**

P.A.$_2$ = % $_{\text{alta o}}$ % $_{\text{media}}$ = %$_1$ + %$_2$ = **44%.**

P.A.$_3$ = % $_{\text{alta o}}$ % $_{\text{media o}}$ % $_{\text{baja}}$ = %$_1$ + %$_2$ + %$_3$ = **100%.**

Esta acumulación de porcentajes lleva por nombre en Estadística, **porcentaje acumulado (P.A.):** es una forma de expresar una proporción o fracción acumulada como una fracción de denominador cien; es decir, el producto de la frecuencia relativa acumulada por cien, solo para variables cuasi – cuantitativas y cuantitativas, se calcula así:

$$P.A. = \frac{N_i}{n} \times 100 = F_i \times 100 .$$

Se representa de la siguiente manera:

Clase social	%	P.A
Alta	16	+ 16
Media	28	+ 44
Baja	56	100
Total	100	No Lleva

La tabla para la variable clase social quedaría del siguiente modo:

Tabla 2. Estudiantes del Colegio Salesiano "San Luis" según la clase social. Mérida – Venezuela. Mayo – 2007.

Clase social	n_i	N_i	f_i	F_i	%	P.A.
Alta	5	5	0,16	0,16	16	16
Media	9	14	0,28	0,44	28	44
Baja	18	32	0,56	1	56	100
Total	32	X	1	x	100	X

Fuente: Coordinación Docente del Colegio Salesiano "San Luis"

Veamos esta tabla con más detenimiento:

La frecuencia absoluta acumulada, frecuencia relativa acumulada y porcentaje acumulado no es necesario que lleven un total, porque el cálculo de la última categoría siempre es igual al total de las frecuencias y porcentajes absolutos.

Tabla 2. Estudiantes del Colegio Salesiano "San Luis" según la clase social. Mérida – Venezuela. Mayo – 2007.

Clase social	ni	Ni	fi	Fi	%	P.A
Alta	5	5	0,16	0,16	16	+ 16
Media	9	+ 14	0,28	+ 0,44	28	44
Baja	18	+ 32	0,56	+ 1	56	+ 100
Total	32	X	1	X	100	X

Fuente: Coordinación Docente del Colegio Salesiano "San Luis"

Siempre son iguales los cuadros coloreados

Representa que no es necesario el total de las frecuencias acumuladas

Problemas de consolidación

1. Construya una tabla con todos sus elementos para la variable nivel de instrucción de la madre e interprete los resultados obtenidos.

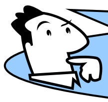
2. Suponga que los siguientes datos fueron recolectados a estudiantes de séptimo grado del Colegio Salesiano "san Luis", Mérida – Venezuela durante el mes de mayo de 2007. Construya una tabla con todos sus elementos para las siguientes variables e interprete los resultados:

Opinión sobre el programa ¿Quién quiere ser millonario?					
Muy bueno	Regular	Malo	Muy malo	Malo	Regular
Bueno	Regular	Bueno	Bueno	Regular	Regular
Bueno	Malo	Bueno	Regular	Malo	Muy bueno
Bueno	Regular	Regular	Bueno	Bueno	Muy malo

Calidad del servicio de cantina.					
Bueno	Malo	Regular	Malo	Regular	Regular
Bueno	Malo	Regular	Bueno	Bueno	Bueno
Regular	Regular	Malo	Regular	Bueno	Regular
Malo	Malo	Malo	Regular	Regular	Regular

Talla de franela					
S	M	M	S	XL	M
M	S	M	L	S	S
M	M	S	M	L	M
M	M	M	S	S	S

<u>Clase 4</u>: Organización de datos para variables cuantitativas o distribución de frecuencias.



El objetivo de la clase es buscar que el alumno utilice variables cuantitativas y organice los datos mediante tablas.

Los datos que se toman como referencia para efectuar las tablas para variables cuantitativas son de los 32 estudiantes del segundo año del ciclo de Educación Media Diversificado tomados al azar del Colegio Salesiano "<u>San Luis</u>" del estado Mérida mostrados en la **TABLA A**.

Se resumen los datos para la variable edad (en meses cumplidos) de la **TABLA A** como en la clase anterior; se sabe que la variable en estudio es una variable cuantitativa discreta porque sus datos son enteros y por tanto se le pueden hallar los diferentes tipos de frecuencia y porcentajes.

<u>La tabla quedará del siguiente modo:</u>

<u>Tabla 3</u>. Estudiantes del Colegio Salesiano "San Luis" según la edad. Mérida – Venezuela. Mayo – 2007.

Edad (meses cumplidos)	n_i	N_i	f_i	F_i	%	P.A.
180	**1**	**1**	**0,03**	**0,03**	**3**	**3**
182	**1**	**2**	**0,03**	**0,06**	**3**	**6**
184	**4**	**6**	**0,13**	**0,19**	**13**	**19**
188	**1**	**7**	**0,03**	**0,22**	**3**	**22**
192	**1**	**8**	**0,03**	**0,25**	**3**	**25**
193	**2**	**10**	**0,06**	**0,31**	**6**	**31**
194	**2**	**12**	**0,06**	**0,37**	**6**	**37**
195	**4**	**16**	**0,13**	**0,50**	**13**	**50**
196	**4**	**20**	**0,13**	**0,63**	**13**	**63**
198	**4**	**24**	**0,13**	**0,76**	**13**	**76**
199	**2**	**26**	**0,06**	**0,82**	**6**	**82**
200	**1**	**27**	**0,03**	**0,85**	**3**	**85**
204	**1**	**28**	**0,03**	**0,88**	**3**	**88**
205	**1**	**29**	**0,03**	**0,91**	**3**	**91**
208	**1**	**30**	**0,03**	**0,94**	**3**	**94**
212	**1**	**31**	**0,03**	**0,97**	**3**	**97**
219	**1**	**32**	**0,03**	**1**	**3**	**100**
Total	32	x	1	X	100	

<u>Fuente:</u> Coordinación Docente del Colegio Salesiano "San Luis".

La tabla 3 es sumamente extensa y no se están resumiendo mucho los datos, entonces lo que se puede inferir a partir de esta tabla, es que para variables cuantitativas se debe hacer otro proceso para resumir estos datos en una tabla.

Un nuevo proceso para agrupar los datos de la tabla 3, ya que la misma no resume mucho los datos, es el siguiente:

- La mayor edad en meses es 219 y la menor es 180; con esto estamos abarcando la totalidad de los datos de la muestra.
- A restar estos valores obtenemos la diferencia de las edades en meses, la cual es 39 meses.
- Y por último le sumamos la unidad a la anterior diferencia, lo que nos da como resultado 40 meses.

Lo que se está haciendo, lleva por nombre, **rango o recorrido (R)**: es la distancia o diferencia que existe entre el máximo y el mínimo valor de una serie de datos más la unidad, se calcula así:

$$R = (ValorMáximo - ValorMínimo) + Unidad.$$

Nota: la unidad depende de cómo estén presentados los datos, **por ejemplo:** si son enteros como los del problema la unidad es 1, si son datos de esta manera: 0.4, 0.7, 1.7, 1.8, 2.3, 2.4, la unidad es 0.1 y si los datos son: 0.42, 1.53, 42.07, la unidad es 0.01 y así sucesivamente.

Lo que se quiere es agrupar los datos de la edad, en cinco (5) números de clases, esto lleva por nombre **el número de intervalos o clases (Nc)**: el número que se tomaran es entre cinco y diez, a criterio personal, ya que un número muy pequeño de clases significaría una agrupación extrema, y en consecuencia sin utilidad; y un número muy grande no nos estaría resolviendo el problema de resumir los datos.

Para obtener la extensión del intervalo de la edad se divide el rango (40) entre el número de clases (5) y se obtiene una extensión de **8**, esto lleva por nombre, **amplitud de la clase o el intervalo (a_i)**: es el cociente del rango de la clase entre el número de clases y se calcula así:

$$a_i = \frac{R}{Nc}.$$

Se debe saber desde que valor inicia y hasta qué valor culmina cada una de las clases; es decir, los límites. Cuando se dice desde qué valor inicia se esta haciendo referencia al extremo de donde comienzan los datos de la clase, en el problema 1, para la variable edad (meses cumplidos) este extremo es 180, y concretamente lleva por nombre **límite aparente inferior (L.A.I.):** es el extremo inferior o valor mínimo del **intervalo**; por tanto, $L.A.I. = ValorMínimo.$ Ahora cuando se dice hasta qué valor culmina, se está haciendo alusión al extremo donde terminan los datos de la clase, por esto le sumamos la amplitud del intervalo y al resultado le restamos la unidad correspondiente, por ende, este extremo es (180+8) - 1=187, y esto lleva por nombre **límite aparente superior (L.A.S.):** es el extremo superior o valor máximo del **intervalo** y se calcula así: $L.A.S. = (ValorMínimo + a_i) + Unidad.$

Con lo anterior se puede construir una **distribución de frecuencias**: este tipo de tablas se realiza para agrupar grandes cantidades de datos de manera conveniente y consiste en especificar intervalos de valores de la característica en estudio, los cuales se denominan intervalos de clase o simplemente clases, para luego determinar las frecuencias y porcentajes en cada una de las clases, este tipo de tablas se realiza solo para variables de tipo cuantitativo.

Veamos como queda la tabla:

Clases	Límite Aparente Inferior (L.A.I.)	Límite Aparente Superior (L.A.S.)
1	180	187
2	188	195
3	196	203
4	204	211
5	212	219

Después de calcular el L.A.I. y L.A.S. de la primera clase, se le suma a_i a cada uno de los límites aparentes.
Esto se realiza con el fin de resumir trabajo

Como se puede observar en la tabla anterior, un valor no pueden estar en dos modalidades o clases a la vez, esta es una condición que debe cumplir la distribución de frecuencias, con esto la clasificación es **mutuamente excluyente**, esto se comprueba observando que el L.A.S. de la primera clase no sea igual al L.A.I de la segunda clase y así sucesivamente; además, todos los sujetos deben estar clasificados, no debe quedar ninguno fuera de la clasificación o distribución, con lo cual la clasificación es **colectivamente exhaustiva**, esto se comprueba observando el último L.A.S. el cual debe ser mayor o igual al valor máximo de la serie de datos.

 A partir de lo antes visto podemos completar la distribución de frecuencias como se presenta a continuación

Tabla 4. Distribución de Frecuencias de la edad (meses cumplidos) de los estudiantes del Colegio Salesiano "San Luis". Mérida – Venezuela. Mayo – 2007.

Clases	L.A.I. – L.A.S.	n_i	N_i	f_i	F_i	%	P.A.
1	180 - 187	6	6	0,19	0,19	19	19
2	188 – 195	10	16	0,31	0,50	31	50
3	196 – 203	11	27	0,35	0,85	35	85
4	204 – 211	3	30	0,09	0,94	9	94
5	212 - 219	2	32	0,06	1	6	100
TOTAL		32	X	1	X	100	X

Fuente: Coordinación Docente del Colegio Salesiano "San Luis"

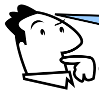
Problemas de consolidación

1. Responda e interprete cada una de las preguntas siguientes considerando la información de la tabla 4:

- ¿Cuántos estudiantes tienen edades comprendidas entre 188 y 211 meses?
- ¿Cuál es la proporción de estudiantes con edades entre 180 y 203 meses?
- ¿Cuál es el porcentaje de estudiantes con edades desde 204 meses hasta 219 meses?
- ¿Cuál es la proporción de estudiantes con edades desde 180 hasta 195 meses y con edades desde 188 hasta 211 meses?
- ¿Cuál es la cantidad de estudiantes con edades desde 180 hasta 187 meses o con edades desde 212 hasta 219 meses?
- ¿Cuál es el porcentaje de estudiantes con edades comprendidas entre 196 y 203 meses o con edades comprendidas entre 212 y 219 meses?

2. Construya una distribución de frecuencias para el promedio aritmético de notas (puntos).

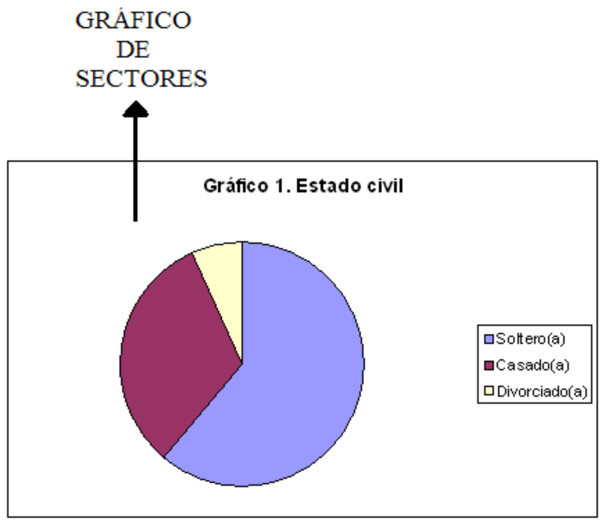
Clase 5: Representación gráfica de las variables cualitativas y cuasi – cuantitativas.



El objetivo de la clase es buscar que los alumnos representen gráficamente las variables cualitativas y cuasi - cuantitativas.

Como su nombre lo indica, los gráficos son una representación más visual y trasmiten en forma inmediata una idea general de los principales aspectos de los datos, hemos visto en las noticias o en los periódicos estos tipos de gráficos, los cuales son muy variados, a continuación se van a estudiar tres tipos de gráficos: gráfico de sectores, gráfico de barras simples e histograma.

GRÁFICO DE SECTORES



En la televisión, el Internet o en el periódico, para diversos temas como por ejemplo las elecciones gubernamentales se observan gráficos como el mostrado anteriormente, que tiene similitud a una torta, recuerden que esa torta se puede dividir en porciones, como si se repartiera en un cumpleaños o en porciones que en el problema 1 de la variable género serían masculino y femenino, a esas porciones se le llaman sectores, y la cantidad de esos sectores son tantos como categorías haya, donde cada sector debe ser proporcional a la cantidad de estudiantes por cada categoría, lleva por nombre **gráficos de sectores:** es de forma circular y solamente lo vamos a utilizar para variables cualitativas, entonces ahora vamos a construir este gráfico:

Se calcula la proporción del sector de la categoría femenino que llamaremos $sector_1$:

$$S_1 = Sector_1 = Femenino = \frac{19}{32} = 0.59.$$

Pero recuerden que se va a tomar un círculo y el mismo esta dividido en 360º, por ello el resultado lo multiplicamos por 360º para obtener los grados correspondientes al sector:

$$S_1 = Sector_1 = 0.59 \times 360^{\circ} = 212.4^{\circ} \approx 212^{\circ}.$$

Ahora se calcula la proporción para la categoría masculino que llamaremos $sector_2$:

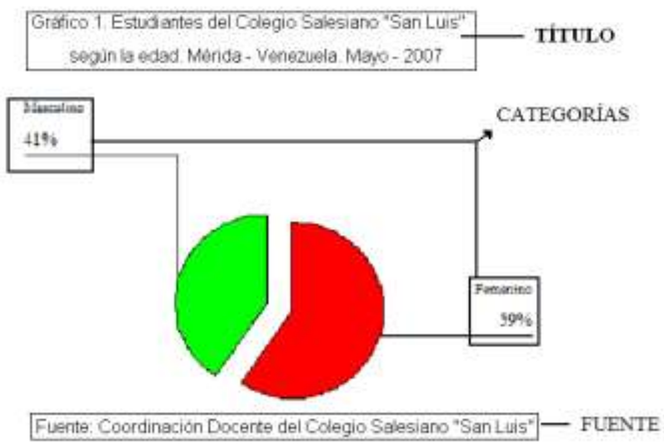
$$S_2 = Sector_2 = Masculino = \frac{13}{32} = 0.41.$$

Luego se multiplica por 360º:

$S_2 = Sector_2 = 0.41 \times 360° = 147.6° \approx 148°$.

Podemos concluir que cada sector se halla de manera general así: $S_i = n_i \times \dfrac{360}{n}$.

Dibujamos un círculo para luego medimos con el transportador los grados correspondientes a cada sector y rellenamos de color y el gráfico quedará del siguiente modo:



Observen que el gráfico tiene título y fuente, también cada sector esta debidamente identificado con la categoría correspondiente.

Interpretación: Se observa que la mayoría de estudiantes son del género femenino con un 59% del total y el 41% de estudiantes restantes son del género masculino.

Problemas de consolidación

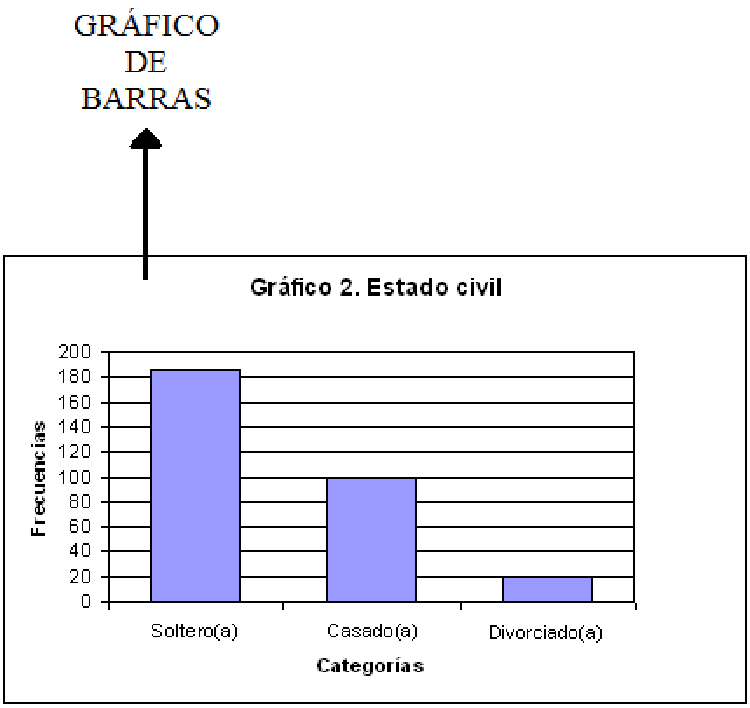
Suponga que los siguientes datos fueron recolectados a estudiantes de séptimo grado del Colegio Salesiano "San Luis", Mérida – Venezuela, durante el mes de mayo de 2007.

1. Construya un gráfico de Sectores con todos sus elementos para cada una de las variables presentadas e interprete los resultados:

Nº	Color de ojos	Opinión sobre el programa ¿Quién quiere ser millonario?
1	Negro	Bueno
2	Marrón	Bueno
3	Gris	Malo
4	Negro	Bueno
5	Negro	Malo
6	Marrón	Bueno
7	Negro	Bueno
8	Gris	Bueno
9	Negro	Malo
10	Gris	Malo

11	Marrón	Bueno
12	Marrón	Malo
13	Marrón	Bueno
14	Negro	Malo
15	Negro	Bueno
16	Negro	Malo
17	Gris	Bueno
18	Negro	Malo
19	Marrón	Bueno
20	Marrón	Malo

GRÁFICO DE BARRAS SIMPLES

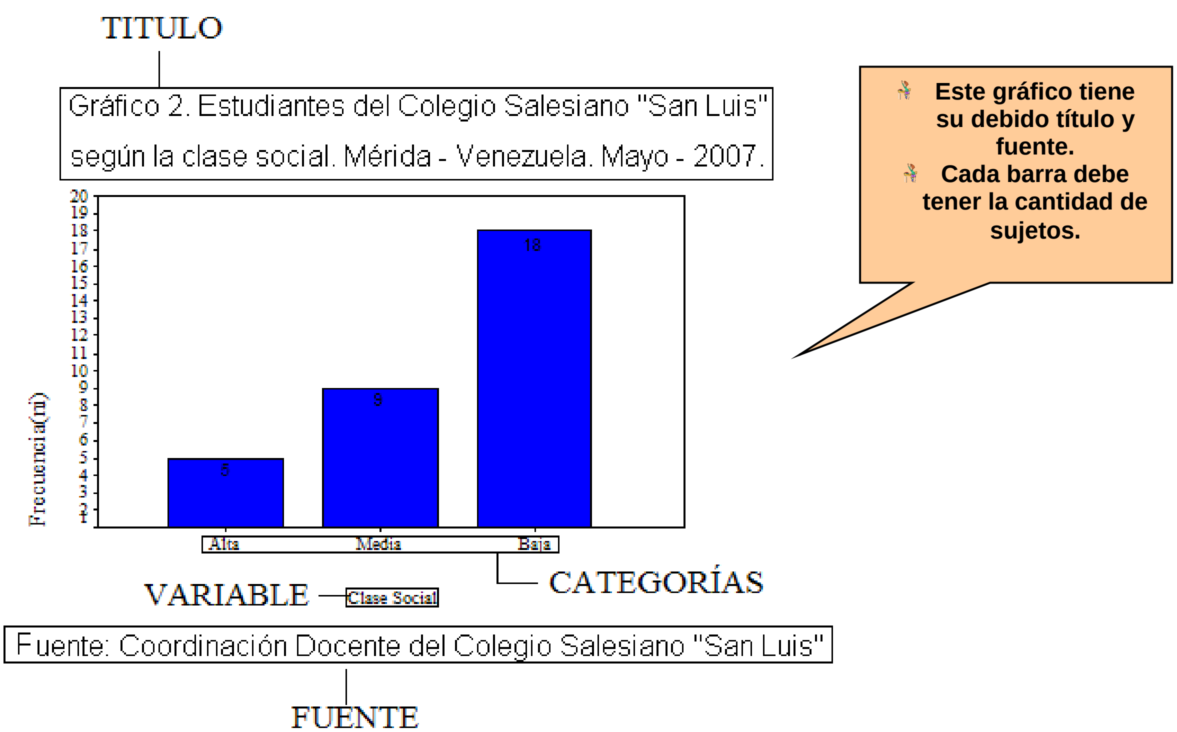


El gráfico anterior tiene forma de barras verticales y cada barra representa una categoría de la variable en estudio, lleva por nombre **gráfico de barras simples:** cada barra forma un rectángulo y simples porque es de una variable, se recomienda para datos cualitativos y cuasi – cuantitativos como por ejemplo clase social y género.

- Para construir un gráfico de barras, se utiliza en el eje de las abscisas o eje “x” las categorías de clase social y en el eje de las ordenadas o eje “y”, el número de datos de la variable o frecuencia absoluta (n_i) de cada categoría, esto es lo primero que haremos en una hoja de papel milimetrado.

- Para cada categoría se levanta una barra en forma rectangular cuya altura viene dada por el número de datos de cada categoría, es lo siguiente que debemos realizar.
- Las barras deben ir separadas y tanto el ancho como la distancia que las separa son arbitrarios, pero una vez fijados deben mantenerse en todo el gráfico por tanto tomemos un ancho de barras y una separación de un centímetro, manteniendo fijas esas medidas.

Se puede ahora construir el gráfico de barras simples, el cual quedará de la manera siguiente:



Interpretación: se puede ver que la mayor cantidad de estudiantes son de clase social baja, seguida por la clase social media y muy pocos estudiantes de clase social alta.

Problemas de consolidación

1. Construya un gráfico de barras simples con todos sus elementos para la variable de instrucción de la madre de la **TABLA A** e interprete los resultados obtenidos.

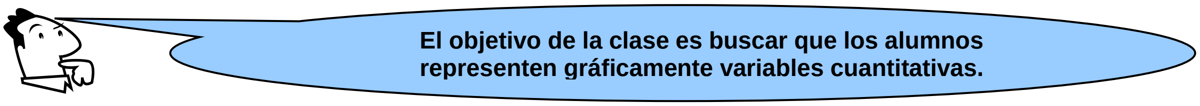
2. Suponga que los siguientes datos fueron recolectados a estudiantes de séptimo grado del Colegio Salesiano "San Luis", Mérida – Venezuela, durante el mes de mayo de 2007. Construya e interprete los resultados de los gráficos de barras simples para las siguientes variables con todos sus elementos:

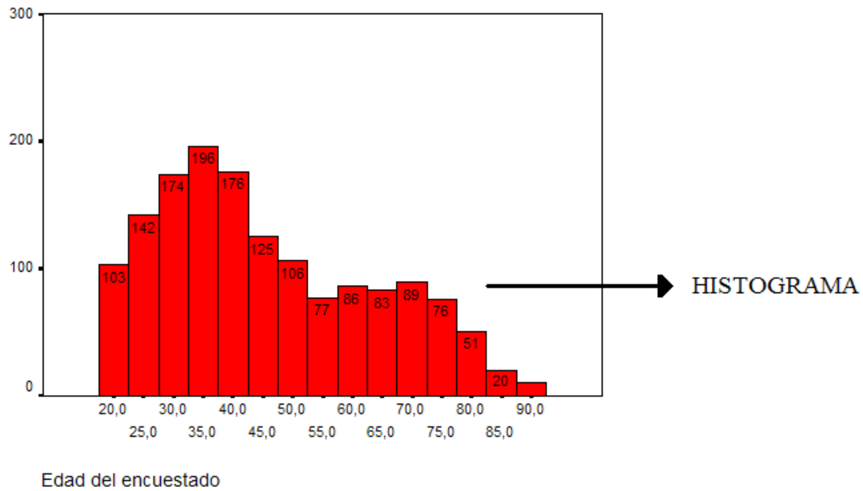
Talla de franela					
S	M	M	S	XL	M
M	S	M	L	S	S
M	M	S	M	L	M
M	M	M	S	S	S

Calidad del servicio de cantina.					
Bueno	Malo	Regular	Malo	Regular	Regular
Bueno	Malo	Regular	Bueno	Bueno	Bueno
Regular	Regular	Malo	Regular	Bueno	Regular
Malo	Malo	Malo	Regular	Regular	Regular

Opinión sobre el programa ¿Quién quiere ser millonario?					
Muy bueno	Regular	Malo	Muy malo	Malo	Regular
Bueno	Regular	Bueno	Bueno	Regular	Regular
Bueno	Malo	Bueno	Regular	Malo	Muy bueno
Bueno	Regular	Regular	Bueno	Bueno	Muy malo

<u>Clase 6:</u> Representación gráfica de variables cuantitativas.





Para representar gráficamente una variable cuantitativa como la edad (meses cumplidos) del problema 1, el gráfico adecuado es el gráfico anterior, que lleva por nombre **<u>histograma:</u>** se utiliza para representar distribuciones de frecuencias cuyas clases son intervalos. Consiste en un gráfico de barras con la particularidad de que las barras están juntas unas de otras.

Para construirlo se utiliza la distribución de frecuencias de la edad

<u>Tabla 4</u>. Distribución de Frecuencias de la edad (meses cumplidos) de los estudiantes del Colegio Salesiano "San Luis". Mérida – Venezuela. Mayo – 2007.

Clases	L.A.I. – L.A.S.	n_i	N_i	f_i	F_i	%	P.A.
1	180 - 187	6	6	0,19	0,19	19	19
2	188 – 195	10	16	0,31	0,50	31	50
3	196 – 203	11	27	0,35	0,85	35	85
4	204 – 211	3	30	0,09	0,94	9	94
5	212 - 219	2	32	0,06	1	6	100
TOTAL		32	X	1	X	100	X

<u>Fuente:</u> Coordinación Docente del Colegio Salesiano "San Luis"

- Para que exista continuidad entre la serie de datos le restamos al límite aparente inferior (L.A.I.) la unidad dividido entre dos y al límite aparente superior (L.A.S.) le sumamos la unidad dividido entre dos, estos límites que son necesarios hallar para nuestra indagación son el **límite real inferior (L.R.I.) y el límite real superior (L.R.S.)** de cada clase o intervalo, que se calculan para asegurar la continuidad de lo datos, se realiza para la primera clase y luego se suma la amplitud del intervalo:

$$L.R.I._1 = L.A.I._1 - \left(\frac{unidad}{2}\right) = 180 - \left(\frac{1}{2}\right) = 179.5.$$

$$L.R.S._1 = L.A.I_1 + \left(\frac{unidad}{2}\right) = 187 + \left(\frac{1}{2}\right) = 187.5.$$

- **Veamos como quedará la tabla 4 con los límites reales:**

En los intervalos que determinan los límites reales siempre el límite real inferior es abierto o excluyente y el límite real superior es cerrado o incluyente.

Tabla 4. Distribución de frecuencias de la edad (meses cumplidos) de los estudiantes del Colegio Salesiano "San Luis". Mérida – Venezuela. Mayo – 2007.

Clases	L.R.I. – L.R.S.	L.A.I. – L.A.S.	n_i	N_i	f_i	F_i	%	P.A.
1	(179.5 – 187.5]	180 – 187	6	6	0,19	0,19	19	19
2	(187.5 – 195.5]	188 – 195	10	16	0,31	0,50	31	50
3	(195.5 – 203.5]	196 – 203	11	27	0,35	0,85	35	85
4	(203.5 – 211.5]	204 – 211	3	30	0,09	0,94	9	94
5	(211.5 – 219.5]	212 - 219	2	32	0,06	1	6	100
TOTAL			32	X	1	X	100	X

Fuente: Coordinación Docente del Colegio Salesiano "San Luis"

- Se construye llevando sobre el eje de abscisas o eje "x" de un plano cartesiano los límites reales de las clases de las edades. Luego en el eje de las ordenadas o eje "y" de un plano cartesiano la cantidad de estudiantes por cada clase.

- Se halla el punto medio de cada intervalo, sumando los limites aparentes de la primera clase y dividimos entre dos, que lleva por nombre **marca de clase (X_m):** se calcula así

$$X_m = \frac{L.A.I. + L.A.S.}{2} = \frac{L.R.I. + L.R.S.}{2}.$$

Para la primera clase la marca de clase será:

$$Xm_1 = \frac{180+187}{2} = \frac{367}{2} = 183.5.$$

Para no realizar este proceso con cada una de las clases, se le suma la amplitud del intervalo, luego de calculada la primera marca de clase.

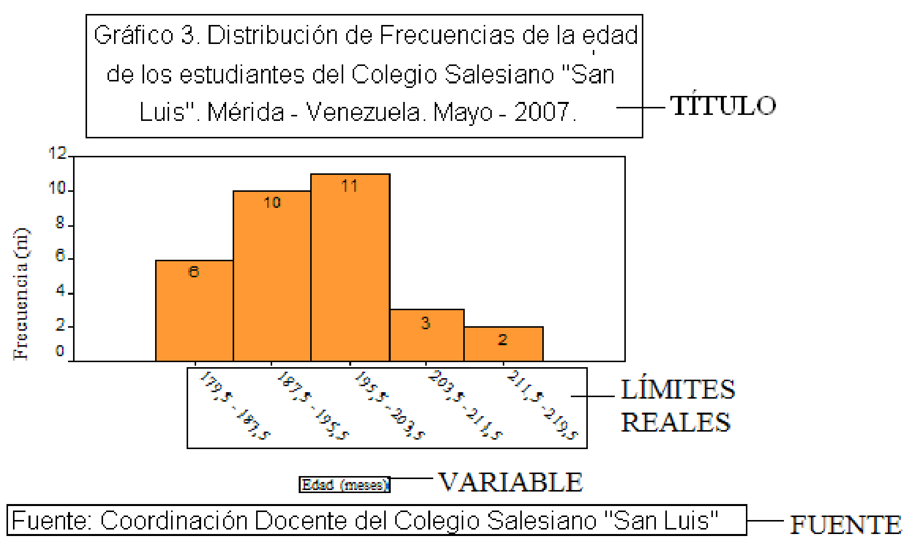
Las marcas de clase quedarían del siguiente modo:

X_m
183.5
191.5
199.5
207.5
215.5

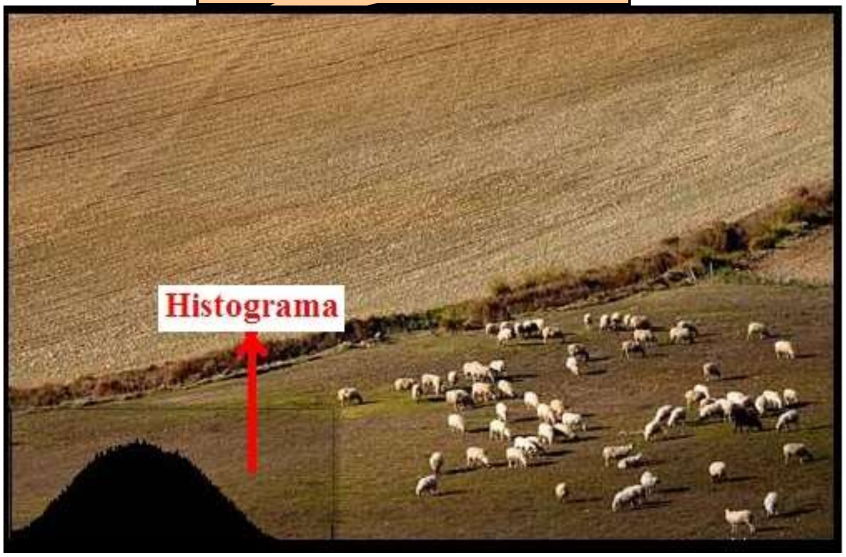
- Se unen los puntos medios o marcas de clase (X_m) con su respectiva frecuencia absoluta, lo cual da la atura de la barra en forma rectangular.
- Las barras deben ir pegadas y el ancho de la misma es arbitraria, pero una vez fijada deben mantenerse en todo el gráfico; para este ejemplo se toma un ancho de barras de un centímetro, manteniendo fijas esas medidas.

Se puede ahora construir un histograma, el cual quedará de la siguiente manera:

Este gráfico posee todos los elementos necesarios, además cada barra lleva sus respectivos límites reales y la cantidad de estudiantes por cada barra.



Aplicación del histograma



En la fotografía es donde tiene aplicación el histograma. Es un gráfico simple que muestra donde se encuentran todos los niveles de luminosidad en la escena, desde las sombras a las altas luces. Estos niveles se disponen de izquierda (sombras) a la derecha (altas luces). La altura que hay en cada nivel muestra cuanta información se encuentra en ese particular nivel.

Además el histograma de esta imagen en concreto se puede apreciar como la cámara divide este gráfico conforme a lo que serían los puntos de diafragma, de modo que si hiciese falta hacer un ajuste en la exposición sería más sencilla; digamos que con el tiempo resultará igual que mirar la hora por un reloj de agujas que aún sin saber los valores exactos sabemos discernir su significado, con el tiempo y una vez que se comprenda el histogramas se puede saber de un vistazo qué calidad tiene la imagen. Esto es especialmente útil cuando se encuentra sobre impuesto en la imagen o a su lado. Sería conveniente aclarar que al igual que no existe una exposición correcta para la película; tampoco existe un histograma ideal que aplicar a todas las imágenes. Excepto cuando las altas luces están completamente quemadas cualquier dibujo de histograma es correcto (siempre y cuando se interprete correctamente).

Problemas de consolidación

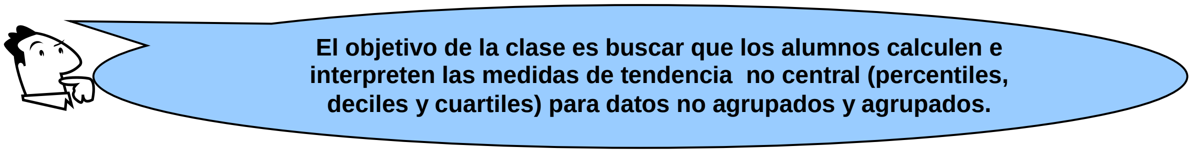
1. Construya un histograma con la distribución de frecuencias del promedio aritmético de notas (puntos) de la **TABLA A**.

2. Suponga que los siguientes datos fueron recolectados a estudiantes de noveno grado del Colegio Salesiano “San Luis”, Mérida – Venezuela, durante el mes de mayo de 2007. Construya un histograma con los siguientes datos:

Estatura (centímetros).		
Clases	**L.A.I. – L.A.S.**	**n_i**
1	155 – 160	8
2	161 – 165	13
3	166 – 170	15
4	171 – 175	7
5	176 – 180	4
6	181 - 185	3
TOTAL		**50**

Tiempo (minutos).		
Clases	**L.A.I. – L.A.S.**	**n_i**
1	30 – 38	15
2	39 – 46	10
3	47 – 54	9
4	55 – 62	6
5	63 – 70	6
6	71 - 78	4
TOTAL		**50**

Clase 7: Medidas de tendencia no central.



Cuando se trate de una cantidad muy numerosa de datos, la condensación y la representación gráfica de los mismos constituyen elementos muy importantes pero no suficientes para realizar un adecuado análisis descriptivo de una colección de ellos. Se hace necesario utilizar ciertos indicadores numéricos, que nos proporcionan una idea más clara sobre aspectos relevantes, estos indicadores se estudiarán con más detalles a continuación.

Cuando se cuantifica un hecho ocurrido en la muestra en estudio, por ejemplo la ubicación del 30% de las edades mayores de los estudiantes se encuentra un valor específico que nos ubica ese 30% de las edades mayores para realizar esto utilizamos las **medidas de tendencia no central:** o también llamadas medidas de posición no central, lleva ese nombre porque se puede tomar un valor que no necesariamente se encuentra en el centro de los datos como en el ejemplo anterior, estas medidas permiten ubicar la posición que ocupa un valor dentro de un conjunto de datos, dentro de ellas se tienen los: percentiles (P_k), deciles (D_k),cuartiles (Q_k).

Ejemplo: se toman los seis primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180, 206.

Calcule e interprete ¿Cuál es la edad que deja el 70% de las edades por debajo?

Se debe calcular el percentil 70 (P_{70})

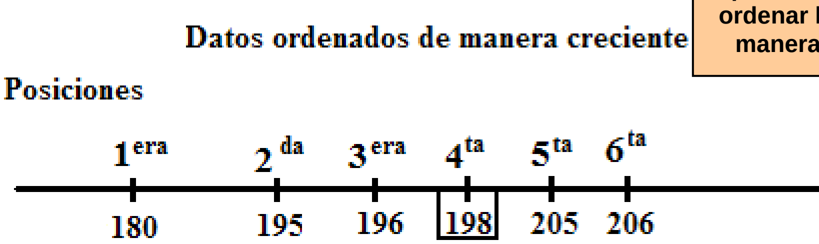
Primero se ordenan los datos de manera creciente:

Datos ordenados: 180, 195, 196, 198, 205, 206.

Luego multiplicamos la cantidad de datos por el percentil a determinar y el resultado se divide entre cien, esto se realiza con el fin de obtener la posición,

$$P_{70} = \frac{6 \times 70}{100} = 4.2 \approx 4.$$

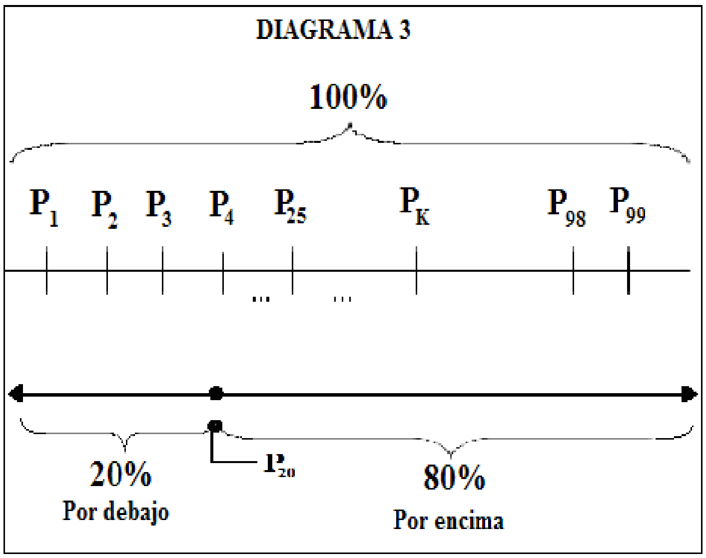
El número calculado indica la posición del percentil, luego de ordenar los datos de manera creciente.



Entonces 4^{ta} es la posición del **P_{70} =198 meses**

Interpretación: 198 meses es la edad que deja el 70% de las edades por debajo y el 30% por encima.

A partir de lo anterior se define a los **percentiles(P_k):** donde k = {1,…,99}, son aquellos valores que dividen los datos ordenados de manera creciente en cien partes iguales. Existen noventa y nueve percentiles, y el subíndice se refiere al porcentaje de casos por debajo del percentil. Entre dos percentiles consecutivos cualesquiera se encuentra un 1% o 1/100 partes de los datos. Por ejemplo, en el percentil 20 (P_{20}), se encuentran el 20% de los datos por debajo y el 80% por encima (como se muestra en el DIAGRAMA 3).



Con este diagrama se explica que los percentiles dividen los datos ordenados en cien partes y un percentil específico deja un porcentaje por encima (80%) y otro por debajo (20%); es decir, a partir de este diagrama inducimos el concepto.

Los percentiles para datos no agrupados se calculan así:

$$P_k = \frac{n \times k}{100}.$$

Siendo:

n : Total de la muestra.
k : Percentil a calcular.

Para su cálculo, debe cumplir que la variable sea de tipo cuasi-cuantitativo o cuantitativo. Los percentiles vienen expresados en las mismas unidades en que vienen expresados los datos.

Ejemplo para datos agrupados: para calcular el percentil 60 (P_{60}) para datos agrupados, se va a tomar la tabla 4,

Primero se comienza con la fórmula de datos no agrupados, para hallar la posición del percentil 60 (P_{60}),

$$P_{60} = \frac{32 \times 60}{100} = 19.2.$$

19.2 Se busca en N_i, no necesariamente debe estar el número mostrado en N_i, lo importante es que ese valor hallado este incluido en la frecuencia acumulada; por tanto 19.2 está incluido en la **tercera clase** de la tabla 4 ya que esta clase incluye los valores desde el 17 hasta el 27, entonces se utiliza el L.R.I de esa clase (195.5).

Se le resta a la posición la frecuencia acumulada anterior ($N_{i(a)}$ =16); a continuación se divide entre la frecuencia absoluta de la clase (n_i=11) y el resultado se multiplica por la amplitud del intervalo (a_i=8) y por último como está ubicada en la tercera clase, se le suma el límite real inferior de esa clase (195.5). Los cálculos quedarían del siguiente modo:

$$P_{60} = 195.5 + \left(\frac{\left(\frac{32 \times 60}{100} \right) - 16}{11} \right) \times 8 = 197.827 \approx 198 \, meses.$$

Interpretación: aproximadamente 198 meses es el valor que deja el 60% de las edades por debajo y el 40% por encima.


Se puede generalizar el cálculo de percentiles para datos agrupados con la siguiente fórmula:

$$P_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{100} \right) - N_{i(a)}}{n_i} \right) \times a_i \, .$$

Siendo:

P_k : Percentil a calcular.
$L.R.I.$: Límite real inferior de la clase del percentil.
n : Número total de datos de la muestra.
$N_{i(a)}$: Frecuencia acumulada anterior a la clase del percentil.
n_i : Frecuencia absoluta de la clase del percentil.
a_i : Amplitud de la clase del percentil.


Ejemplo: se toman los seis primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180, 206.

Calcule e interprete ¿Cuál es la edad que deja el 20% de las edades por encima?

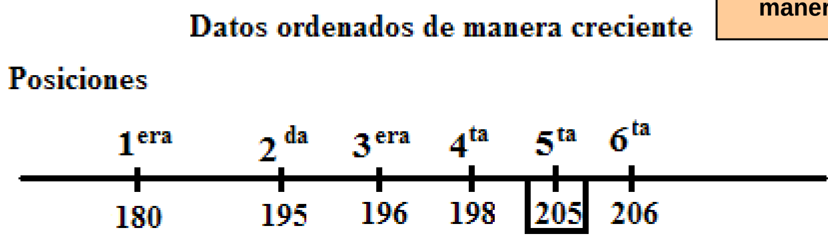
Primero se ordenan los datos de manera creciente:

Datos ordenados: 180, 195, 196, 198, 205, 206.

Luego se multiplica la cantidad de datos por el decil a determinar y el resultado se divide entre diez, esto se realiza con el fin de obtener la posición,

$$D_8 = \frac{6 \times 8}{10} = 4{,}8 \approx 5.$$

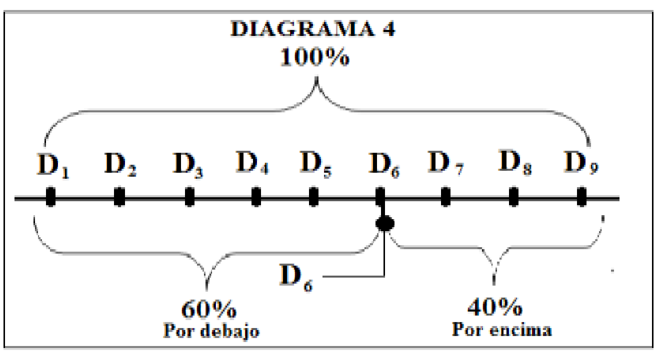
El número calculado indica la posición del decil, luego de ordenar los datos de manera creciente.



Entonces 5^{ta} es la posición del D_8 **=205 meses**

Interpretación: 205 meses es el valor que deja el 80% de las edades por debajo y el 20% por encima.

Entonces los **deciles (D_k):** donde k = {1,…,9}, son aquellos valores que dividen los datos ordenados en diez partes iguales. Existen nueve, el subíndice se refiere al porcentaje de casos por debajo del decil. Entre dos deciles consecutivos cualesquiera se encuentra un 10% o 1/10 partes de los datos .Por ejemplo en el decil (D_6) se encuentran el 60% de los datos por debajo y el 40% por encima (como se muestra en el DIAGRAMA 4).



Con este diagrama se explica que los deciles dividen los datos ordenados en diez partes y un decil específico deja un porcentaje por encima (40%) y otro por debajo (60%), es decir, a partir de este diagrama inducimos el concepto.

De manera general los deciles, se calculan así para datos no agrupados: $D_k = \frac{n \times k}{10}$.

Para su cálculo, debe cumplir que la variable sea de tipo cuasi-cuantitativo o cuantitativo. Los deciles vienen expresados en las mismas unidades en que vienen expresados los datos.

Siendo:

n : Total de la muestra.

k : Decil a calcular.

Ejemplo para datos agrupados: para calcular el decil 8 (D_8) para datos agrupados, se van a tomar los datos de la tabla 4.

Primero se comienza con la formula de datos no agrupados, para hallar la posición del decil 8 (D_8),

$$D_8 = \frac{32 \times 8}{10} = 25.6.$$

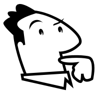
25,6 se busca en N_i, no necesariamente debe estar el número mostrado en N_i, lo importante es que ese valor hallado este incluido en la frecuencia acumulada; por tanto 25,6 está incluido en la **tercera clase** de la tabla 4 ya que esta clase incluye los valores desde el 17 hasta el 27, entonces se utiliza el L.R.I de esa clase (195.5).

Se le resta a la posición la frecuencia acumulada anterior ($N_{i(a)}$ =16); a continuación se divide entre la frecuencia absoluta de la clase (n_i=11) y el resultado se multiplica por la amplitud del intervalo (a_i=8) y por último como está ubicada en la tercera clase, debemos sumar el límite real inferior de esa clase (195.5).

Los cálculos quedarían del siguiente modo:

$$D_8 = 195.5 + \left(\frac{\left(\frac{32 \times 6}{10} \right) - 16}{11} \right) \times 8 = 202.48 \approx 203 \, meses.$$

Interpretación: aproximadamente 203 meses es el valor que deja el 80% de las edades por debajo y el 20% por encima.


Se puede generalizar el cálculo de percentiles para datos agrupados con la siguiente fórmula:

$$D_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{10}\right) - N_{i(a)}}{n_i} \right) \times a_i \,.$$

Siendo:

D_k : Decil a calcular.
$L.R.I.$: Límite real inferior de la clase del decil.
n : Número total de datos de la muestra.
$N_{i(a)}$: Frecuencia acumulada anterior a la clase del decil.
n_i: Frecuencia absoluta de la clase del decil.
a_i: Amplitud de la clase del decil.


Ejemplo: se toman los seis primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180, 206.

Calcule e interprete ¿Cuál es el valor que toma como máximo el 25% de las edades de los estudiantes?

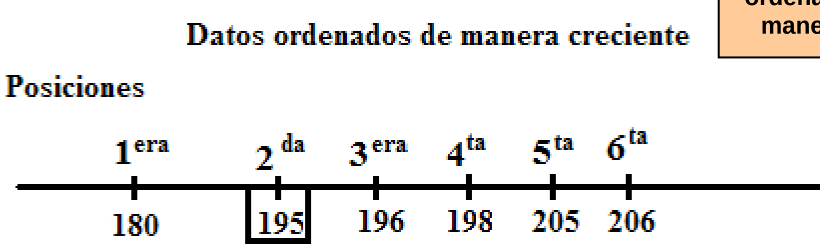
Primero se ordenan los datos de manera creciente:

Datos ordenados: 180, 195, 196, 198, 205, 206.

Luego se multiplica la cantidad de datos por el cuartil a determinar y el resultado se divide entre cuatro, esto se realiza con el fin de obtener la posición,

$$Q_3 = \frac{6 \times 1}{4} = 1{,}5 \approx 2.$$

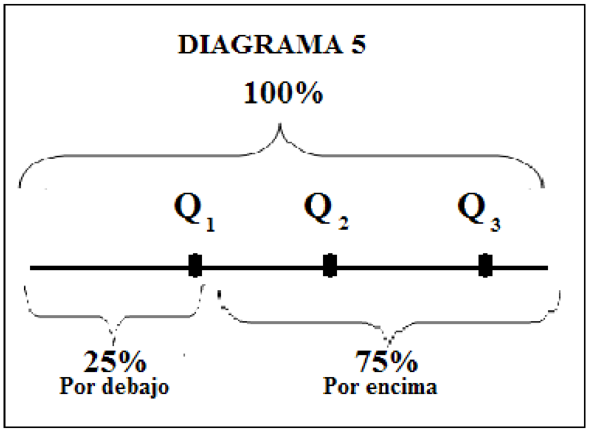
El número calculado indica la posición del cuartil, luego de ordenar los datos de manera creciente.



Entonces 2da es la posición del Q_1 **=195 meses.**

Interpretación: 195 meses es la edad máxima del 25% de los estudiantes.

Por consiguiente, los **cuartiles (Q_k):** donde k = {1, 2, 3} son aquellos valores que dividen los datos ordenados en cuatro partes iguales. Existen tres, el subíndice se refiere al porcentaje de casos por debajo del cuartil. Entre dos cuartiles consecutivos cualesquiera se encuentra un 25% o 1/4 partes de los datos .Por ejemplo en el cuartil (Q_1) se encuentran el 25% de los datos por debajo y el 75% por encima.



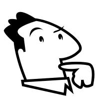
Con este diagrama se explica que los cuartiles dividen los datos ordenados en cuatro partes y un cuartil específico deja un porcentaje por encima (75%) y otro por debajo (25%), es decir, a partir de este diagrama se induce el concepto.

De manera general los cuartiles, se calculan así para datos no agrupados: $Q_k = \frac{n \times k}{4}$.

Para su cálculo, debe cumplir que la variable sea de tipo cuasi-cuantitativo o cuantitativo. Los cuartiles vienen expresados en las mismas unidades en que vienen expresados los datos.

Siendo:

n : Total de la muestra.
k : Cuartil a calcular.

**Ejemplo para datos agrupados:** para calcular el cuartil 1 (Q_1) para datos agrupados, vamos a tomar la tabla 4.

Se comienza con la fórmula de datos no agrupados, para hallar la posición del cuartil 1 (Q_1),

$$Q_1 = \frac{32 \times 1}{4} = 8.$$

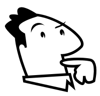
8 se busca en N_i, no necesariamente debe estar el número mostrado en N_i, lo importante es que ese valor hallado este incluido en la frecuencia acumulada; por tanto 8 está incluido en la **segunda clase** de la tabla 4 ya que esta clase incluye los valores desde el 7 hasta el 16, así se utiliza el L.R.I. de esa clase (187.5).

Se le resta a la posición la frecuencia acumulada anterior ($N_{i(a)}$ =6); a continuación se divide entre la frecuencia absoluta de la clase (n_i=10) y el resultado se multiplica por la amplitud del intervalo (a_i=8) y por último como está ubicada en la segunda clase, debemos sumar el límite real inferior de esa clase (187.5).

Los cálculos quedarían del siguiente modo:

$$Q_1 = 187.5 + \left(\frac{\left(\frac{32 \times 1}{4} \right) - 6}{10} \right) \times 8 = 189.1 \approx 189 \, meses.$$

Interpretación: aproximadamente 189 meses es el valor que toma como máximo el 25% de las edades de los estudiantes.

Se puede generalizar el cálculo de cuartiles para datos agrupados con la siguiente fórmula:

$$Q_k = L.R.I. + \left(\frac{\left(\frac{n \times k}{4} \right) - N_{i(a)}}{n_i} \right) \times a_i \, .$$

Siendo:

Q_k : Cuartil a calcular.

$L.R.I.$: Límite real inferior de la clase del cuartil.

n : Número total de datos de la muestra.

$N_{i(a)}$: Frecuencia acumulada anterior a la clase del cuartil.

n_i : Frecuencia absoluta de la clase del cuartil.

a_i : Amplitud de la clase del cuartil.

Problemas de consolidación

1. Calcular e interpretar el P_{35}, D_2, Q_1, Utilizando los nueve primeros datos de la variable promedio aritmético de notas (puntos) de la **TABLA A.**

2. Calcular e interpretar el P_{15}, D_7, Q_3, Utilizando los nueve primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

3. Utilizar la distribución del promedio aritmético de notas (puntos) de la **TABLA A** para desarrollar las siguientes preguntas y elabore su debida interpretación.

- ¿Por debajo de que valor se encuentra el 10% de la distribución del promedio de notas de los estudiantes?
- ¿Por encima de que valor se encuentra el 90% de la distribución del promedio de notas de los estudiantes?
- ¿Cuál sería el puntaje mínimo del promedio de notas que un estudiante debería tener para estar ubicado en el 25% de los puntajes más altos?
- ¿Cuál sería el puntaje máximo del promedio de notas que un estudiante debería tener para estar ubicado en el 25% de los puntajes más bajos?

4. Utilizar la distribución de la edad (meses cumplidos) de la **TABLA A** para desarrollar las siguientes preguntas y elabore su debida interpretación.

- ¿Por debajo de que valor se encuentra el 20% de la distribución del promedio de notas de los estudiantes?
- ¿Por encima de que valor se encuentra el 70% de la distribución del promedio de notas de los estudiantes?
- ¿Cuál sería el puntaje mínimo del promedio de notas que un estudiante debería tener para estar ubicado en el 25% de los puntajes más altos?
- ¿Cuál sería el puntaje máximo del promedio de notas que un estudiante debería tener para estar ubicado en el 25% de los puntajes más bajos?

<u>Clase 8:</u> Medidas de tendencia central.

El objetivo de la clase es buscar que los alumnos calculen e interpreten las medidas de tendencia central: moda, mediana y media.

Cuando se desea información sobre los valores medios de la serie de datos de la muestra, se utilizan una serie de medidas que tratan de representar o resumir los datos obtenidos, además estas sirven para realizar comparaciones entre datos de muestras diferentes, estas son llamadas **<u>medidas de tendencia central:</u>** son las que indican la posición hacia la que tienden a concentrarse las observaciones o alrededor del cual se distribuyen el conjunto de datos.

<u>¿Qué está de moda?</u>

- Los celulares, las computadoras, RBD, el Internet, ver televisión, la política, jugar fútbol, entre otros.

Todo lo antes mencionado es lo que está de moda y se sabe **<u>¿por qué está de moda?</u>**

- La mayoría de las personas quieren comprar un celular, una computadora, un cd de un artista actual.
- De igual forma la mayoría de las personas quieren navegar en Internet un par de horas, o distraerse jugando fútbol un rato, o en su defecto hablar un rato de política.

Entonces, en una serie de datos se define **<u>moda (M_o):</u>** aquel valor que se repite más, es decir que se repite con más frecuencias. La moda es el valor más común entre los datos y viene expresada en las mismas unidades de los datos.

<u>Ejemplo:</u> se toman los siete últimos datos de la variable edad (meses cumplidos) de la **TABLA A.**

<u>Datos</u>: 196**,** 199, 198, 207, 204, 201, 196.

Calcule e interprete ¿<u>Cuál es el valor que más se repite</u>?

Es la única medida de posición que se puede hallar para los diferentes tipos de variables. En los datos es posible tener una o más de una moda e incluso es posible que no exista. Esto último ocurre cuando no hay un valor que se repita más que los demás.

Para datos no agrupados

Debemos ver cuál es el valor que más se repite y esa es la **<u>moda</u> (Mo)**.

Mo=196 meses.

<u>Interpretación:</u> 196 meses es el valor que más se repite o de mayor frecuencia.

Se puede generalizar que la moda (Mo) para datos no agrupados se determina, contando cuál es el valor que más se repite.

Ejemplo para datos agrupados: para calcular la **moda (Mo)** para datos agrupados, se va a tomar la tabla 4.

Clases	L.R.I. – L.R.S.	n_i	
1	(179.5 – 187.5]	6	
2	(187.5 – 195.5]	10	Frecuencia anterior
3	(195.5 – 203.5]	11	Frecuencia modal
4	(203.5 – 211.5]	3	Frecuencia posterior
5	(211.5 – 219.5]	2	
TOTAL		32	

- Se comienza buscando la clase con más cantidad (n_i) de estudiantes del mismo modo como se realiza para datos no agrupados, para hallar la posición de la moda o clase modal.
- La posición de la moda (Mo) está en la tercera clase que es llamada la **frecuencia modal**, porque su frecuencia absoluta es la de mayor cantidad; por lo tanto esa es la clase modal y el L.R.I a usar es 195.5.
- A continuación, se obtiene (d_1) que es la resta de la frecuencia modal (11) menos la **frecuencia anterior** (10) y el resultado es **1.**
- Luego, se obtiene (d_2) que es la resta de la frecuencia modal (11) menos la **frecuencia posterior** (3) y el resultado es **8**.
- Esto se realiza para tomar el valor modal central, por ello se realiza el resultado de d_1 y d_2 por separado.
- d_1 se divide entre $d_1 + d_2$, el resultado se multiplica por la amplitud del intervalo (a_i=8) y por último como está ubicada en la tercera clase, debemos sumar el límite real inferior de esa clase (195.5).

Los cálculos quedan del siguiente modo:

$$Mo = 195.5 + \left(\frac{1}{1+8}\right) \times 8 = 196{,}38 \approx 196\,meses.$$

Interpretación: aproximadamente 196 meses es la edad que más se repite o de mayor frecuencia.

De manera general la moda (Mo), se calcula así para datos agrupados:

$$Mo = L.R.I. + \left(\frac{d_1}{d_1 + d_2} \right) \times a_i .$$

Siendo:

$L.R.I.$: Límite real inferior de la clase modal.

a_i: Amplitud de la clase modal.

$d_1 = Frecuencia \quad \text{mod}\,al - \quad Frecuencia \quad anterior\,.$

$d_2 = Frecuencia \quad \text{mod}\,al - \quad Frecuencia \quad posterior\,.$

Ejemplo: se toman los cinco primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180.

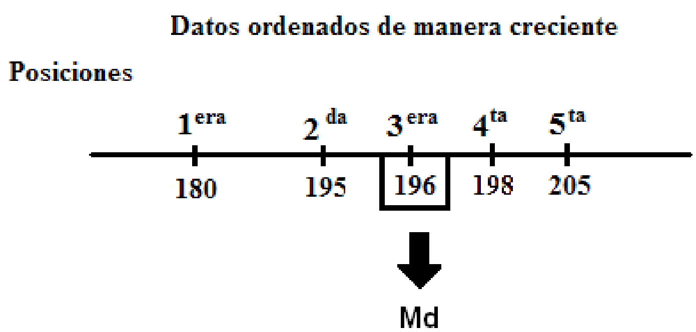
Calcule e interprete

¿Cuál es el valor que ocupa la posición central?

Primero se ordenan los datos de manera creciente:

Datos ordenados: 180, 195, 196, 198, 205.

La mediana viene expresada en las mismas unidades en que vienen los datos y no tiene porque ser igual a uno de los valores existentes en el grupo de datos. Se puede utilizar para variables cuasi-cuantitativas y cuantitativas.



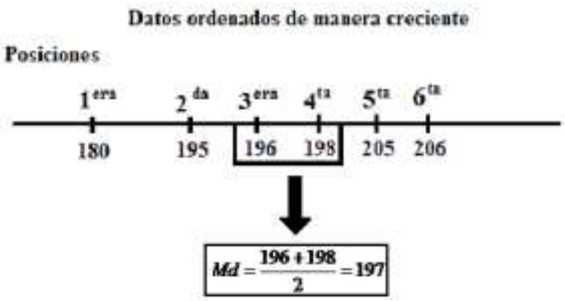
Interpretación: 196 meses es el valor central de las distintas edades en estudio.

Nota: Para datos no agrupados, cuando el número de datos es impar, hay un solo valor que ocupa la posición central y en consecuencia ese valor es la mediana.

Ejemplo: se toman los seis primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180, 206.

Calcule e interprete ¿Cuál es el valor que deja el 50% de los datos por debajo y el 50% por encima?



Interpretación: 197 meses es el valor que deja el 50% de los datos por debajo y el 50% por encima.

Mientras que cuando el número de datos es par, existen dos valores que ocupan la posición central, entonces se toma como mediana el promedio entre éstos.

El valor central de los datos en estudio, es llamado **mediana(Md):** de una compilación de datos, que previamente han sido ordenados en forma creciente o decreciente, se define como aquel valor que a lo sumo es menor que la mitad de los datos y que a su vez es mayor a lo sumo, que la mitad de los datos. Alternativamente podemos decir, que la mediana es aquel valor que ocupa la posición central de los datos, una vez que estos han sido ordenados. En forma aproximada podemos decir que el 50% de los datos están por debajo de la mediana y el 50% por encima.

Ejemplo para datos agrupados: para calcular la **mediana (Md)** para datos agrupados, se toma la tabla 4.

Clases	L.R.I. – L.R.S.	n_i	N_i
1	(179.5 – 187.5]	6	6
2	(187.5 – 195.5]	10	16
3	(195.5 – 203.5]	11	27
4	(203.5 – 211.5]	3	30
5	(211.5 – 219.5]	2	32
	TOTAL	32	X

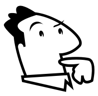
CLASE ANTERIOR

CLASE MEDIANAL

FRECUENCIA DE LA CLASE MEDIANAL

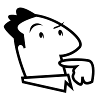
- Se comienza dividiendo el total de datos de la muestra entre dos, esto determina la posición de la mediana (Md) o clase medianal.
- La posición de la mediana (Md) se ubica en la frecuencia acumulada, por tanto Md está en la segunda clase que es llamada la **clase medianal**, porque la posición esta incluida en esta clase y el L.R.I a usar es 187.5.
- A continuación, se realiza la resta de la posición de la clase medianal (16) menos la **frecuencia acumulada de la clase anterior** (6) el resultado es 10**.**
- Esto se realiza para tomar el valor central, por ello se le resta la frecuencia acumulada anterior.

- Se divide entre 10 que es la frecuencia de la clase medianal, el resultado (1) se multiplica por la amplitud del intervalo (a_i=8) y por ultimo como está ubicada en la tercera clase, debemos sumar el límite real inferior de esa clase (187.5).

Los cálculos quedarían del siguiente modo:

- $$Md = 187.5 + \left(\frac{16 \quad - \quad 6}{10} \right) \times 8 = 195{,}5 \approx 196\,meses.$$

<u>Interpretación</u>: aproximadamente 196 meses es la edad que divide la distribución en dos partes iguales.

De manera general la mediana (Md), se calcula así para datos agrupados:

$$Md = L.R.I + \left(\frac{n/_2 - N_{i(a)}}{n_i} \right) \times a_i .$$

**<u>Ejemplo:</u>** se toman los siete últimos datos de la variable edad (meses cumplidos) de la **TABLA A.**

<u>Datos</u>: 196**,** 199, 198, 207, 204, 201, 196.

Calcule e interprete <u>¿Cuál es el promedio de las edades</u>?

Para datos no agrupados

La media aritmética viene expresada en las mismas unidades en que vienen expresados los datos, se interpreta como el valor promedio de una serie de datos y se calcula solo para variables cuantitativas.

Para determinar cuál es el promedio de las edades se calcula la **<u>media</u> (** $\bar{x}$ **)**.

$$\bar{x} = \frac{196 + 199 + 198 + 207 + 204 + 201 + 196}{7} = \frac{1401}{7} = 200{,}143 \approx 200\,meses.$$

$\bar{x}$ **=200 meses.**

<u>Interpretación:</u> aproximadamente 200 meses es el promedio de las edades.

Cuando se termina cualquier materia y se calcula un promedio, se suman las notas y se dividen entre la cantidad de notas, esto se llama **<u>media</u>**

aritmética ($\overline{x}$): de una colección de datos $x_1,\ldots, x_n$ se define como la suma de esos datos, dividida entre el número de sumandos.

De manera general la media aritmética ($\overline{x}$), se calcula así para datos no agrupados:

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

Ejemplo para datos agrupados:

para calcular la media **(** $\bar{x}$ **)** para datos agrupados, se toman los datos de la tabla 4, y las marcas de clase (X_m), utilizadas para realizar el histograma.

L.A.I.-L.A.S.	n	X_m	$n \times X_m$
180-187	6	183.5	1101
188-195	10	191.5	1915
196-203	11	199.5	2194.5
204-211	3	207.5	622.5
212-219	2	215.5	431
TOTAL	32	X	6234

$\Rightarrow \sum n_i \times X_m$

- Se multiplica cada marca de clase con su correspondiente frecuencia absoluta (n_i), luego el resultado de cada producto se suma, y por último se divide entre el total de datos de la muestra, tal como se hace a continuación: $\bar{x} = \dfrac{6264}{32} = 195{,}75 \approx 196\,meses.$

Interpretación: aproximadamente 196 meses es el promedio de la edad de los estudiantes.

De manera general la media aritmética ($\overline{x}$), se calcula así para datos agrupados:

$$\bar{x} = \frac{X_{m_1} + \ldots + X_{m_k}}{n_1 + \ldots + n_k} = \frac{\sum_{i=1}^{k} X_{m_i} n_i}{\sum_{i=1}^{k} n_i}, donde \quad k = Nc.$$

Problemas de consolidación

1. Calcular e interpretar las medidas de tendencia central para los siguientes datos de estaturas en metros (m.) tomados al azar:

Datos: 1,76; 1,70; 1,69; 1,60; 1,65; 1,70; 1,73; 1,66; 1,80.

2. Responda para la variable promedio aritmético de notas (puntos) de la TABLA A las siguientes preguntas:

- ¿Cuál es el valor central de los datos?
- ¿Cuál es el promedio de la sección?
- ¿Cuál es el valor de mayor frecuencia?

3. Calcular las medidas de tendencia central para cada una de las siguientes variables (Nota: tome en consideración el tipo de variable).

Peso (kg.)	Cargo en una institución Educativa	Estado Civil
80.23	Director (a)	Casado (a)
75.32	Profesor (a)	Soltero (a)
66.4	Profesor (a)	Soltero (a)
69	Obrero (a)	Soltero (a)
79,1	Obrero (a)	Casado (a)
79,3	Obrero (a)	Soltero (a)
69	Subdirector (a)	Divorciado(a)
78,45	Obrero (a)	Soltero (a)

<u>Clase 9:</u> Medidas de dispersión absolutas.

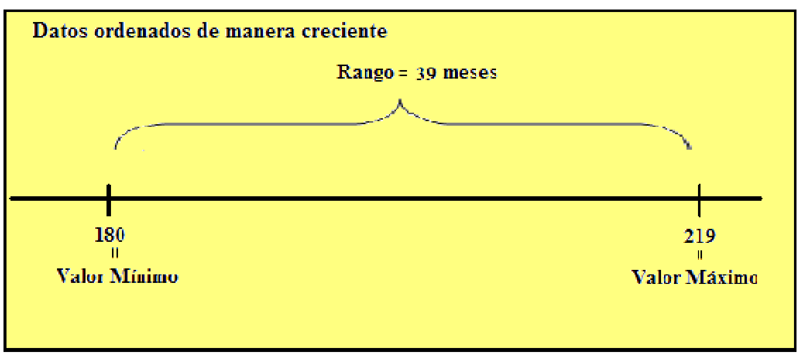
El objetivo de la clase es buscar que los alumnos calculen e interpreten los diferentes tipos de rangos.

Las medidas de dispersión absolutas también llamadas de variabilidad absolutas permiten cuantificar o expresar la variación o el grado de distanciamiento de un grupo de valores respecto a un valor medio de la serie de datos, proporcionando una idea clara sobre cómo se encuentran distribuidos los datos; es decir, si se encuentran muy cercanos o alejados; además el resultado del cálculo de estas medidas, vienen expresadas en la misma unidad de medida de los datos en estudio y son solo para variables cuantitativas.

Cuando se hace la diferencia entre valores extremos de cualquier serie de datos cuantitativos, estamos realizando un **rango**, veamos los principales:

<u>Ejemplo:</u> calcular e interpreta la diferencia entre los valores extremos, tomando en cuenta, las distribución de datos de la variable edad (meses cumplidos) de la **TABLA A.**

Rango= 219-180=39meses.



<u>Interpretación:</u> 39 meses es la diferencia que existe entre mayor y el menor valor de los datos de la variable edad.

Este rango es el más usado, simplemente llamado **<u>rango(R):</u>** es la diferencia entre los valores extremos de una distribución de datos y es la medida de variabilidad más sencilla.

Se generaliza el cálculo del rango (R) para datos no agrupados y agrupados con la siguiente fórmula:

$$R = (ValorMáximo - ValorMínimo).$$

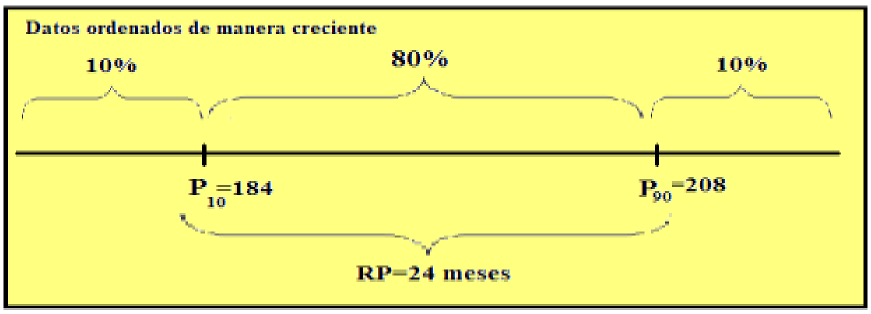
<u>Ejemplo:</u> usando la distribución de datos de la variable edad (meses cumplidos) de la TABLA A, obtenga el 80% de los datos centrales.

Para esto se toma los resultados del cálculo del percentil noventa (P_{90}) y el percentil diez (P_{10}), (se manipulan los datos de los problemas de consolidación de la clase 7) y se realiza su diferencia:

P_{90}=208.3≈208 meses.

P_{10}=183.76≈184 meses.

Rango Percentil (RP)= 208 – 184 = 24 meses.



<u>Interpretación:</u> 24 meses es la dispersión del 80% de los datos de las edades o los datos se encuentran en el intervalo (184; 208).

Por tanto, el **<u>rango percentílico (RP):</u>** es una medida de dispersión basada en percentiles, es la diferencia entre el percentil noventa (P_{90}) y el percentil diez (P_{10}). Esta medida también puede expresarse equivalentemente como un intervalo de la forma (P_{10}; P_{90}). Este rango o intervalo incluye el 80% de los datos de los datos ubicados en la parte central de la distribución; es decir, excluye el 10% inferior y el 10% superior de los datos.

Se generaliza el cálculo del rango percentílico (RP) para datos no agrupados y agrupados con la siguiente fórmula:

$$RP = P_{90} - P_{10}.$$

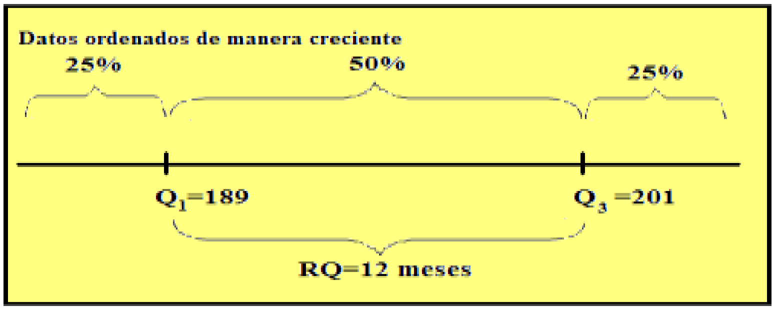
<u>Ejemplo:</u> utilizando la distribución de datos de la variable edad (meses cumplidos) de la **TABLA A**, se busca obtener el 50% de los datos centrales.

Para esto se toman los resultados del cálculo del percentil veinticinco (P_{25}=Q_1) y el percentil setenta y cinco (P_{75}=Q_3), (se manejan los datos de los problemas de consolidación de la clase 7) y se realiza su diferencia:

P_{25}=Q_1=189.1≈189 meses.

P_{75}=Q_3=201.318≈201 meses.

Rango cuartil=201 – 189 = 12 meses.



Interpretación: 12 meses es la dispersión del 50% de los datos de las edades o los datos se encuentran en el intervalo (189; 201).

Entonces, el **Rango cuartílico (RQ):** es una medida de dispersión basada en percentiles, es la diferencia entre los percentiles veinticinco ($P_{25}=Q_1$) y setenta y cinco ($P_{75}=Q_3$). Esta medida también puede expresarse equivalentemente como un intervalo de la forma (Q_1; Q_3). Este rango o intervalo incluye el 50% de los datos de los datos ubicados en la parte central de la distribución; es decir, excluye el 25% inferior y el 25% superior de los datos.

Se generaliza el cálculo del Rango cuartílico (RQ) para datos no agrupados y agrupados con la siguiente fórmula:

$$RQ = Q_3 - Q_1.$$

Problemas de consolidación

1. Calcular e interpretar los diferentes tipos de rangos para los siguientes datos de estaturas en metros (m.) y Pesos en kilogramos (Kg.) de nueve estudiantes tomados al azar:

Datos de estaturas (m.): 1,76; 1,70; 1,69; 1,60; 1,65; 1,70; 1,73; 1,66; 1,80.
Datos de Pesos (Kg.): 80,23; 75,32; 66.4; 69; 79,1; 79,3; 69; 78,45; 77,5.

2. Para la variable promedio aritmético de notas (puntos) de la TABLA A, responda las siguientes preguntas:

- ¿Cuál es la diferencia entre los valores extremos?
- ¿Entre qué valores se encuentra el 80% de los datos centrales?
- ¿Cuál es el intervalo que cubre el 50% de los datos?
- ¿Cuál es el grado de dispersión de los datos?

<u>Clase 10:</u> Medidas de dispersión absolutas alrededor de la media.

El objetivo de la clase es buscar que los alumnos calculen e interpreten las medidas de dispersión muéstrales y poblacionales y la regla empírica de una curva normal o aproximadamente.

Las medidas de dispersión que definen a una población se denominan **<u>parámetros</u>**: es un término utilizado muy frecuentemente en el lenguaje común y no necesariamente con demasiada propiedad. Parámetro es condición variable a la que se asignan unos valores determinados y fijos a partir del resultado obtenido de una muestra para luego limitarlo a una población, resumida para su estudio. Se considera como un valor verdadero de la característica estudiada, para obtener conclusiones racionales y eficaces.

Generalmente los parámetros nos son desconocidos por lo que se deben **<u>"estimar"</u>** a partir de las medidas de dispersión de las muestras que se denomina **<u>estimación puntual</u>**: consiste en utilizar el valor de un estadístico para calcular el parámetro de una población. Por ejemplo, a partir de la media de la edad se calcula la media poblacional (μ) pues a partir de ellos se estiman los parámetros de la población. Los parámetros se suelen identificar con letras griegas y los estimadores con letras latinas.

La dispersión mide que tan alejados están un conjunto de valores; es decir, determina si están muy cercanos o separados, por tanto menor dispersión implica que los datos están más cerca del valor promedio o media aritmética ($\overline{x}$), este aspecto es de vital importancia para el estudio de este tema.

<u>Ejemplo:</u> se toman los siete últimos datos de la variable edad (meses cumplidos) de la **TABLA A.**

<u>Datos</u>: 196**,** 199, 198, 207, 204, 201, 196.

Calcule ¿<u>Cuál es la desviación cuadrática de los valores respecto a la media aritmética</u>?

$$\bar{x} = \frac{196+199+198+207+204+201+196}{7} = \frac{1401}{7} = 200{,}143 \approx 200\,meses.$$

Para calcular la varianza, es recomendado realizar la siguiente tabla:

Xi	Xi - $\bar{x}$	(Xi - $\bar{x}$)^2
196	196 – 200 = - 4	(-4)^2 = 16
199	199 – 200 = - 1	(-1)^2 = 1
198	198 – 200 = - 2	(-2)^2 = 4
207	207 – 200 = 7	(7)^2 = 49
204	204 – 200 = 4	(4)^2 = 16
201	201 – 200 = 1	(1)^2 = 1
196	196 – 200 = - 4	(-4)^2 = 16
		103

$\rightarrow \sum(X_i - \bar{X})^2$

$$Varianza = \frac{103}{7-1} = \frac{103}{6} = 17{,}17 meses^2.$$

<u>Interpretación:</u>

No posee interpretación.

A partir de lo anterior se define la **<u>varianza muestral (S^2):</u>** la media aritmética de las desviaciones cuadráticas de los valores de la variable respecto a su propia media aritmética y se expresa en unidades cuadráticas.

Se generaliza el cálculo de la varianza (S^2) para datos no agrupados con la siguiente fórmula:

$$S^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n-1}.$$

Se divide entre n -1 para obtener una cantidad que permite introducir una corrección matemática en los cálculos estadísticos y para restricciones impuestas en los datos a esto se le denominada <u>grados de libertad.</u> La varianza se calcula con relación al promedio de los datos, para el cálculo de la varianza se le debe restar al total de la muestra el valor uno.

<u>La varianza poblacional (σ^2)</u>, toma en cuenta todos los sujetos de la población y se estima a partir de la varianza muestral sin tomar en cuenta los grados de libertad, entonces para el ejemplo en estudio $\sigma^2 = \frac{103}{7} = 14{,}71 meses^2$.

Se generaliza el cálculo de la varianza poblacional (σ^2) para datos no agrupados con la siguiente fórmula:

$$\sigma^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n}.$$

<u>Ejemplo:</u> Se sigue utilizando la distribución de datos de la variable edad (meses cumplidos) de la **TABLA A**, para obtener las desviaciones cuadráticas de los valores de la variable respecto a su propia media aritmética.

Para calcular la varianza para datos agrupados se utiliza la siguiente tabla:

ni	Xm - $\bar{x}$	(Xm - $\bar{x}$)^2	(Xm - $\bar{x}$)^{2}x ni
6	183.5- 196= -12.5	(-12.5)^2= 156.25	156.25x6= 937.5
10	191.5- 196= -4.5	(-4.5)^2= 20.25	20.25x10= 202.5
11	199.5- 196= 3.5	(3.5)^2= 12.25	12.25x11= 134.75
3	207.5- 196= 11.5	(11.5)^2= 132.25	132.25x3= 396.75
2	215.5- 196= 19.5	(19.5)^2= 380.25	180.25x2= 760.5
			2075

$\Rightarrow \sum (X_m - \bar{X})^2 \times n_i$

$$S^2 = \frac{2075}{32-1} = \frac{2075}{31} = 66{,}9355 \quad meses^2.$$

$$\sigma^2 = \frac{2075}{32} = 64{,}84375 \quad meses^2.$$

Interpretación:

No posee interpretación.

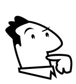
Se generaliza el cálculo de la varianza muestral (S^2) para datos agrupados con la siguiente fórmula:

$$S^2 = \frac{\sum \left(X_m - \bar{x} \right)^2 \times n_i}{n-1}.$$

Se generaliza el cálculo de la varianza poblacional (σ^2) para datos no agrupados con la siguiente fórmula:

$$\sigma^2 = \frac{\sum \left(X_m - \bar{x} \right)^2 \times n_i}{n}.$$

Se utilizan los resultados obtenidos en la varianza muestral (S^2), para obtener una nueva medida de dispersión que posee interpretación, esta medida se llama, **desviación típica muestral o desviación estándar muestral (S):** es la raíz cuadrada de la media aritmética de las desviaciones cuadráticas de los valores de la variable, con respecto a su propia media aritmética de la muestra y viene expresada en unidades originales. Cuando se toma en cuenta población se denomina **desviación típica poblacional o desviación estándar poblacional (σ):** es la raíz cuadrada de la media aritmética de las desviaciones cuadráticas de los valores de la variable, con respecto a su propia media aritmética de la población.

 Entonces utilizando el resultado de la varianza muestral de datos no agrupados, la desviación típica muestral (S), sería la raíz cuadrada de la varianza muestral (S^2); es decir:

$S = \sqrt{17.17 meses^2} = 4{,}134 meses.$ **Interpretación:** aproximadamente 4,134 meses es el monto (o grado) de variabilidad (o dispersión) de los datos muéstrales tomados de la edad.

Para la desviación estándar poblacional, se utiliza el resultado de la varianza poblacional de datos no agrupados, la desviación típica poblacional (σ), sería la raíz cuadrada de la varianza poblacional (σ^2); es decir:

$\sigma = \sqrt{14.71 meses^2} = 3{,}835 \approx 3{,}84 meses.$ **Interpretación:** aproximadamente 3,84 meses es el monto (o grado) de variabilidad (o dispersión) de los datos poblacionales de la edad.

Ahora utilizando el resultado de la varianza muestral de datos agrupados, la desviación típica muestral (S), sería la raíz cuadrada de la varianza muestral (S^2); es decir:

$S = \sqrt{66.9355 \quad meses^2} = 8.18 meses.$ **Interpretación:** aproximadamente 8,18 meses es el monto de variabilidad o dispersión del total de datos de la muestra.

Se aprovecha el resultado de la varianza poblacional (σ^2) de datos agrupados, la desviación típica poblacional (σ), sería la raíz cuadrada de la varianza poblacional (σ^2); es decir:

$$\sigma = \sqrt{64.84375 \quad meses^2} = 8.05 meses.$$

Se generaliza el cálculo de la desviación típica muestral (S) para datos no agrupados y agrupados con la siguiente fórmula:

$$S = \sqrt{S^2}.$$

Se generaliza el cálculo de la desviación típica poblacional (σ) para datos no agrupados y agrupados con la siguiente fórmula:

$$\sigma = \sqrt{\sigma^2}.$$

Regla empírica de la distribución normal o distribución de Gauss

- Media, mediana y moda coinciden.

En el cálculo para datos agrupados de la variable edad, la Moda = Mediana = Media = 196 meses.

Lleva el nombre de regla empírica porque está basada en la experiencia y está sujeta a comprobación y comparación de los datos en estudio con los valores de probabilidad teóricos.

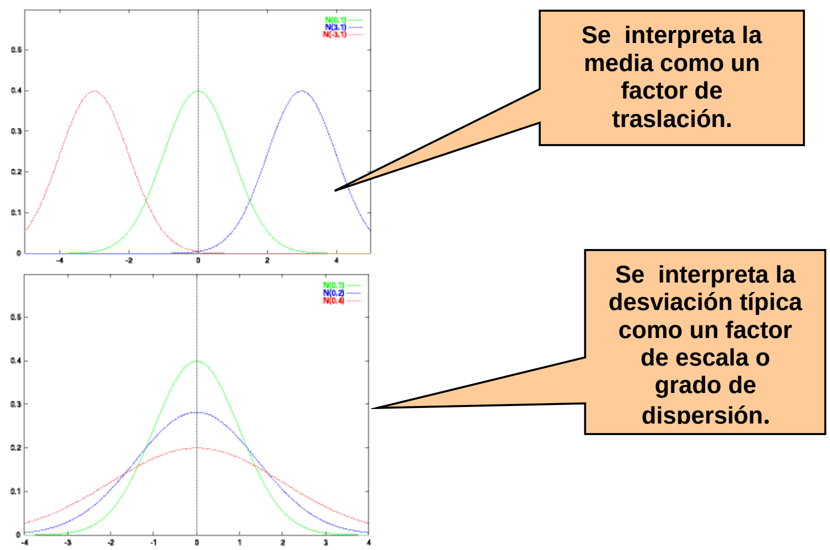
- Es uní modal (posee una sola moda). La Moda es única = 196.
- **<u>Está caracterizada por dos parámetros:</u>** La media poblacional (μ) y la desviación típica poblacional (σ).

$$\overline{X} \approx \mu = 196 meses.$$

$$\sigma = 8.05 meses.$$

N (196,8.05).

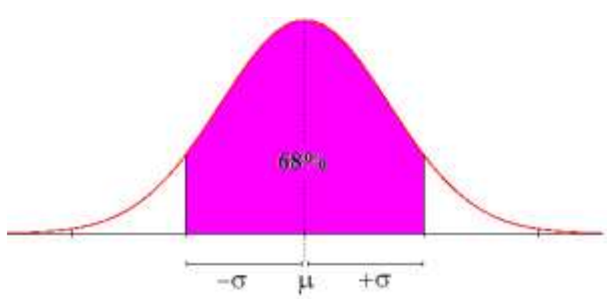
N (μ, σ): Interpretación geométrica.



N (μ, σ): interpretación probabilista.

- Entre la media y una desviación típica se encuentran aproximadamente el 68% de probabilidad:

Si se toma intervalos centrados en μ, y cuyos extremos están a distancia **σ** ➔ se tiene que la probabilidad es del **68%.**

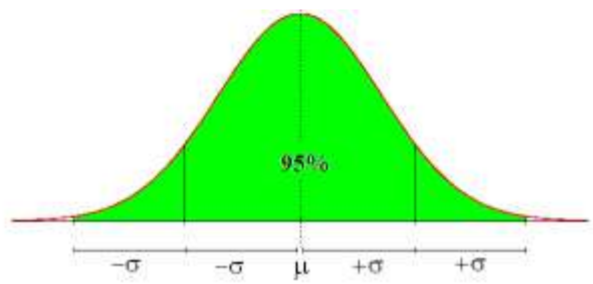


Ejemplo: considerando la distribución de las edades entonces entre 196 ± 8.05 están aproximadamente el 68% de los datos (de manera teórica).

Entre [187.95, 204.05] hay 20 datos lo que representa el 62,5% de probabilidad de que un dato de la variable edad en meses cumplidos sea tomado al azar se encuentre en el intervalo antes mencionado (de manera empírica).

- Entre la media y dos desviaciones típicas se encuentran aproximadamente el 95% de probabilidad:

  Si se toma intervalos centrados en μ, y cuyos extremos están a distancia 2**σ** ➔ se tiene que la probabilidad es del **95%.**

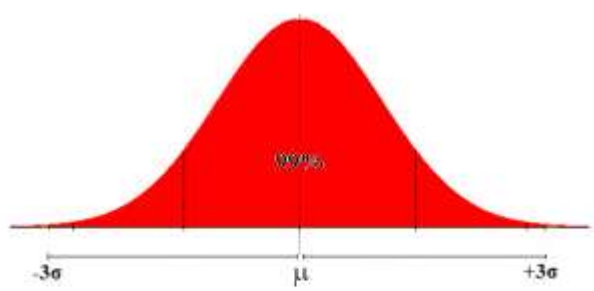


Ejemplo: considerando la distribución de las edades entonces entre $196 \pm 2(8.05)$ están aproximadamente el 95% de los datos (de manera teórica).

Entre [179.9, 212.1] hay 30 datos lo que representa el 93,75% de probabilidad de que un dato de la variable edad en meses cumplidos sea tomado al azar se encuentre en el intervalo antes mencionado (de manera empírica).

- Entre la media y dos desviaciones típicas se encuentran aproximadamente el 99% de probabilidad:

  Si se toma intervalos centrados en μ, y cuyos extremos están a distancia 3**σ** ➔ se tiene que la probabilidad es del **99%.**



Ejemplo: considerando la distribución de las edades entonces entre $196 \pm 3(8.05)$ están aproximadamente el 99% de los datos (de manera teórica).

Entre [171.85, 220.15] existen 32 datos lo que representa el 100% de probabilidad de que un dato de la variable edad en meses cumplidos sea tomado al azar se encuentre en el intervalo antes mencionado (de manera empírica).

Problemas de consolidación

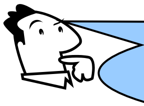
1. Calcular e interpretar los diferentes tipos de medidas de dispersión para los siguientes datos de estaturas en metros (m.) y Pesos en kilogramos (Kg.) de nueve estudiantes tomados al azar:

Datos de estaturas (m.): 1,76; 1,70; 1,69; 1,60; 1,65; 1,70; 1,73; 1,66; 1,80.
Datos de Pesos (Kg.): 80,23; 75,32; 66.4; 69; 79,1; 79,3; 69; 78,45; 77,5.

2. Para la variable promedio aritmético de notas (puntos) de la TABLA A responda las siguientes preguntas:

- ¿Cuál es el grado de dispersión de los datos de la muestra?
- ¿Cuál es el grado de variabilidad de los datos de la población?
- Si cumple con las condiciones de la regla empírica de una Distribución Normal o de Gauss ¿Cuál es la probabilidad de que un sujeto tomado al azar se encuentre en el intervalo que comprende $\bar{x} \pm 2\sigma$?

Clase 11: Medidas de dispersión absolutas basadas en un valor central determinado



El objetivo de la clase es buscar que los alumnos calculen e interpreten las medidas de dispersión: desviación media, desviación mediana.

Están constituidas por aquellas medidas que miden la variabilidad de los datos con respecto a un valor central definido y en consecuencia están basadas en las distancias o desvíos entre cada uno de los datos y el valor central seleccionado; vienen expresadas en la misma unidad de medida de los datos en estudio.

Ejemplo: se toman los siete últimos datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 196**,** 199, 198, 207, 204, 201, 196.

Calcule e interprete ¿Cómo se desvían en promedio los datos de la variable edad?

Se debe calcular la **desviación media** de datos no agrupados, para su cálculo es recomendado realizar la siguiente tabla:

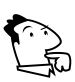
X_i	$X_i - \bar{x}$	$\lvert X_i - \bar{x} \rvert$
196	196 – 200 = - 4	$\lvert -4 \rvert = 4$
199	199 – 200 = - 1	$\lvert -1 \rvert = 1$
198	198 – 200 = - 2	$\lvert -2 \rvert = 2$
207	207 – 200 = 7	$\lvert 7 \rvert = 7$
204	204 – 200 = 4	$\lvert 4 \rvert = 4$
201	201 – 200 = 1	$\lvert 1 \rvert = 1$
196	196 – 200 = 4	$\lvert -4 \rvert = 4$
		23 → $\sum \lvert x_i - \bar{x} \rvert$

$\bar{x}$ **= 200 meses.**

$$Desviación \quad Media = \frac{23}{7} = 3.285 \approx 3.29 meses.$$

Interpretación: los datos tomados al azar de la variable edad se desvían en promedio aproximadamente 3.29 meses.

En consecuencia se define la **desviación media (DM)**: es la media aritmética de los desvíos o diferencias en valor absoluto, de cada uno de los datos con respecto a su media.



Se generaliza el cálculo de la desviación media (DM) para datos no agrupados con la siguiente fórmula:

$$DM = \frac{\sum_{i=1}^{n} \left| X_i - \bar{X} \right|}{n}.$$

Ejemplo para datos agrupados: para calcular la **desviación media (DM)** para datos agrupados, se toma como referencia la tabla 4: $\bar{x}$ **= 196 meses.**

Para calcular esto se utiliza la siguiente tabla:

ni	Xm - $\bar{x}$	|Xm - $\bar{x}$|	|Xm - $\bar{x}$| x ni
6	183.5- 196= -12.5	|-12.5| = 12.5	12.5x6= 75
10	191.5- 196= -4.5	|-4.5| = 4.5	4.5x10= 45
11	199.5- 196= 3.5	|3.5| = 3.5	3.5x11= 38.5
3	207.5- 196= 11.5	|11.5| = 11.5	11.5x3= 34.5
2	215.5- 196= 19.5	|19.5| = 19.5	19.5x2= 39
			232

$\rightarrow \sum |X_m - \bar{X}| \times n_i$

$$DM = \frac{232}{32} = 7.25 meses.$$

Interpretación: los datos de la variable edad (de la TABLA A) se desvían en promedio 7.25 meses.

Se generaliza el cálculo de la desviación media (DM) para datos agrupados con la siguiente fórmula:

$$DM = \frac{\sum_{i=1}^{k} |X_m - \bar{X}| \times n_i}{n}, donde \quad k = Nc.$$

Ejemplo: se toman los cinco primeros datos de la variable edad (meses cumplidos) de la **TABLA A.**

Datos: 205, 196, 195, 198, 180.

Determine ¿Cómo se desvían con respecto al valor central los datos de las edades?

Se debe calcular la **desviación mediana** de datos no agrupados, para su cálculo es recomendado realizar la siguiente tabla:

Xi	Xi - Md	|Xi - Md|
205	205 – 196 = 9	|9| = 9
196	196 – 196 = 0	|0| = 0
195	195 – 196 = - 1	|-1| = 1
198	198 – 196 = 3	|3| = 3
180	180 – 196 = - 16	|-16| = 16
		29

$\rightarrow \sum |X_i - Md|$

Md = 196 meses.

$$Desviación \quad Mediana = \frac{29}{5} = 5.8 meses.$$

Interpretación: los datos tomados al azar de la variable edad se desvían con respecto al valor central en 5.8 meses.

Entonces se define la **desviación mediana (DMd)**: es el valor central de los desvíos o diferencias en valor absoluto, de cada uno de los datos con respecto a su mediana.

Se generaliza el cálculo de la desviación mediana (DMd) para datos no agrupados con la siguiente fórmula:

$$DMd = \frac{\sum_{i=1}^{n} \left| X_i - \overline{X} \right|}{n}.$$

<u>Ejemplo para datos agrupados:</u> para calcular la **<u>desviación mediana (DMd)</u>** para datos agrupados, se toma como referencia la tabla 4, **Md = 195.5 meses.**

<u>Para calcular esto se emplea la siguiente tabla:</u>

ni	Xm - Md	|Xm - Md|	|Xm - Md| x ni
6	183.5- 195.5= -12	|-12| = 12	12x6= 72
10	191.5- 195.5= -4	|-4| = 4	4x10= 40
11	199.5- 195.5= 4	|4| = 4	4x11= 44
3	207.5- 195.5= 12	|12| = 12	12x3= 36
2	215.5- 195.5= 20	|20| = 20	20x2= 40
			232 $\Rightarrow \sum \left|Xm - Md\right| \times n_i$

$$DMd = \frac{232}{32} = 7.25 meses.$$

<u>Interpretación:</u> los datos de la variable edad (de la TABLA A) se desvían con respecto a la mediana 7.25 meses.

Se generaliza el cálculo de la desviación mediana (DMd) para datos agrupados con la siguiente fórmula:

$$DM = \frac{\sum_{i=1}^{k} \left| X_m - Md \right| \times n_i}{n}, donde \quad k = Nc.$$

<u>Problemas de consolidación</u>

1. Calcular e interpretar las Medidas de Dispersión estudiadas para los siguientes datos de estaturas en metros (m.) y Pesos en kilogramos (Kg.) de nueve estudiantes tomados al azar:

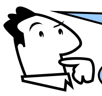
<u>Datos de estaturas (m.):</u> 1,76; 1,70; 1,69; 1,60; 1,65; 1,70; 1,73; 1,66; 1,80.

<u>Datos de Pesos (Kg.):</u> 80,23; 75,32; 66.4; 69; 79,1; 79,3; 69; 78,45; 77,5.

2. Responda las siguientes preguntas:

- ¿Cómo se desvían en promedio los datos en estudio del promedio aritmético de notas (puntos) de la TABLA A?
- ¿Cómo se desvían con respecto al valor central de los datos en estudio los datos en estudio del promedio aritmético de notas (puntos) de la TABLA A?

<u>Clase 12:</u> Excel.



El objetivo de la clase es utilizar las herramientas estadísticas de la hoja de calculo EXCEL para que los alumnos calculen e interpreten los contenidos aprendidos en las clases anteriores.

Hace un tiempo atrás se utilizaban software o programas que permitían crear de una manera fácil planillas o cuadros estadísticos, estos programas fueron evolucionando hasta llegar a lo que hoy conocemos como **Excel.**

<u>Microsoft Office</u>: es un paquete que tiene diferentes aplicaciones como el Word, Excel, PowerPoint, Access una de las primeras versiones fue Office 4.2 luego aparecieron el Office 95, Office 97, Office 2000, Office 2003 y Office 2007. Para nuestro estudio se empleará el Microsoft Office 2003.

<u>Microsoft Excel:</u> es una hoja electrónica que nos permite construir planillas, cuadros estadísticos, registros de asistencias de notas o cualquier serie de datos.

Para ingresar o ejecutar el programa se debe hacer clic en:

DIAGRAMA 6

El Botón Inicio.

Todos los programas

Microsoft Office.

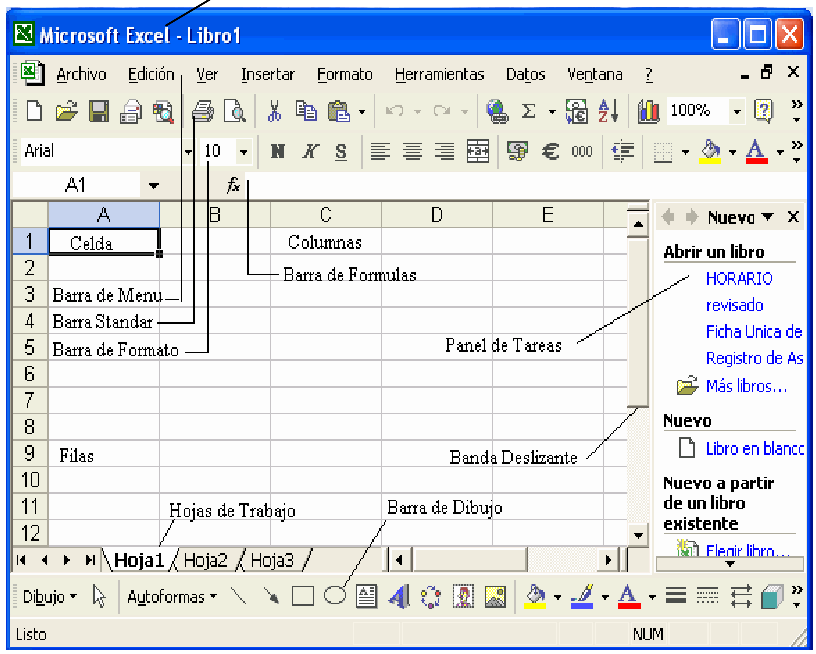
Excel 2003 tal como se muestra en el **DIAGRAMA 6.**

Nos mostrará la ventana de bienvenida de Microsoft Office Excel 2003.

Reconocimiento de la hoja Electrónica

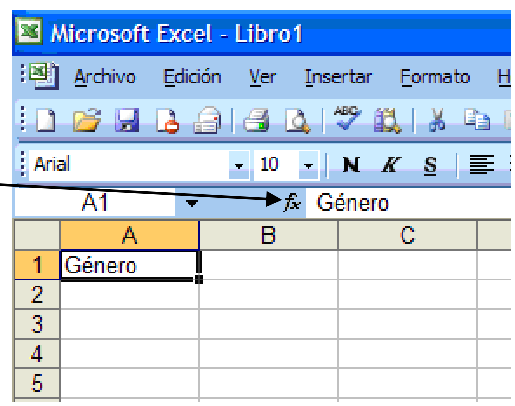
Barra de Título

- Consta de 65536 Filas.
- Y Las columnas están en forma de letras de A hasta IV.
- Contiene Celdas cada una de ellas son separadas.
- Se pueden elaborar y grabar varias hojas de trabajo en un mismo archivo.



Una vez ingresado se debe reconocer las herramientas con las cuales se pueden realizar hojas de cálculo con los datos del problema 1.

<u>**Por ejemplo:**</u> anotar los datos de la variable Género, la celda activa se muestra recuadrada en negrilla, en la celda activa **A1** escribimos delante del signo "f_x" **"Género"** y pulsamos "Enter" en el teclado. Luego anotar cada uno de los datos debajo de la palabra género.



En la primera fila de cada columna siempre se van a anotar el nombre de las variables en estudio debajo del nombre los respectivos datos, se repite el procedimiento para el caso de la variable "edad (meses cumplidos)" .

<u>**Problemas de consolidación**</u>

1. Repita el procedimiento para las variables Clase social, Nivel de instrucción y Promedio de notas aritmético (puntos).

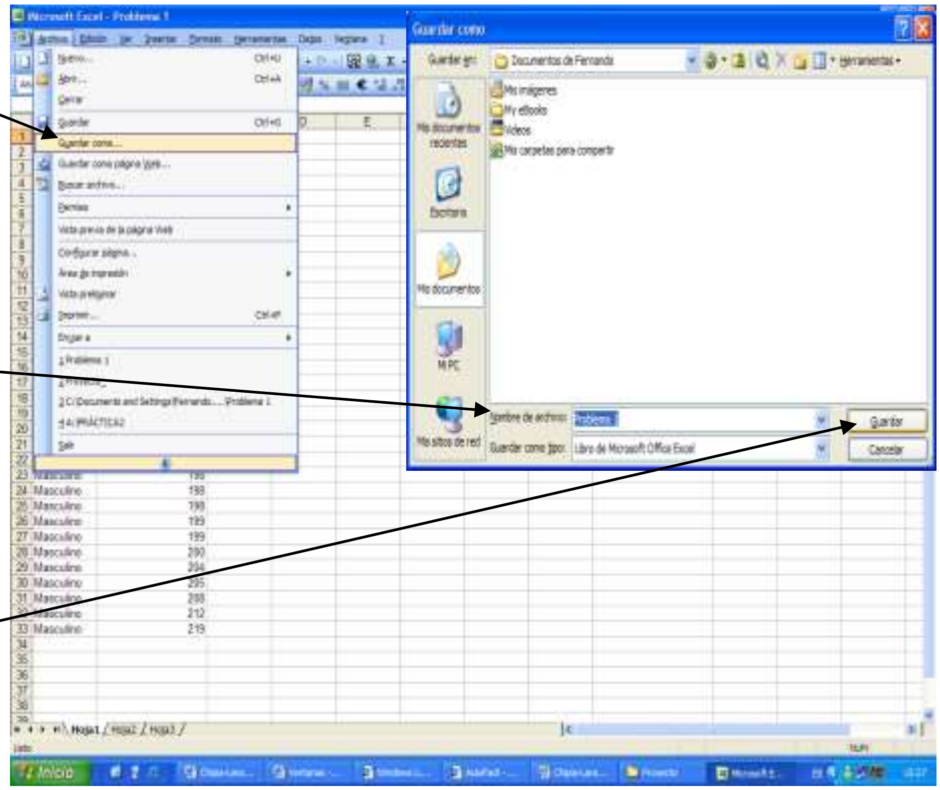
<u>Nota:</u> se debe utilizar la coma (,) para los valore con decimales.

Al terminar se va a guardar los datos

En la Barra de Menú se hace clic sobre **Archivo** y **Guardar como…**

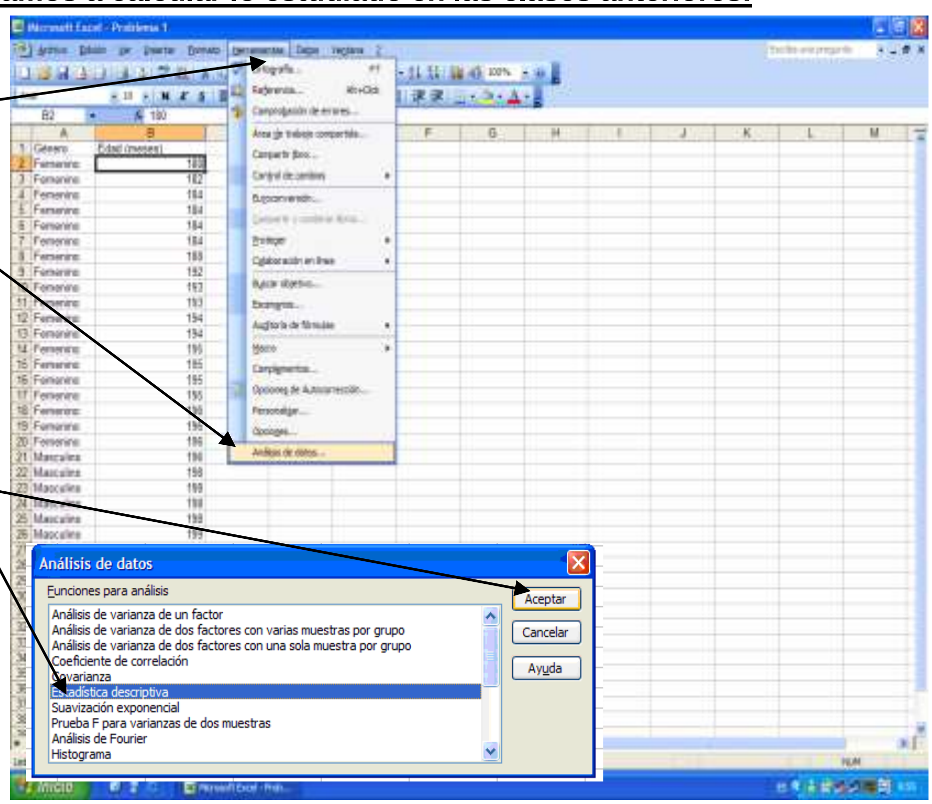
En el formulario que aparece se le coloca un nombre al archivo que guardamos. En este caso se le llama “Problema 1”. Es aconsejable ser ordenado y asignar diferentes directorios a los distintos tipos de archivos y trabajos.

Y clic en **Guardar**

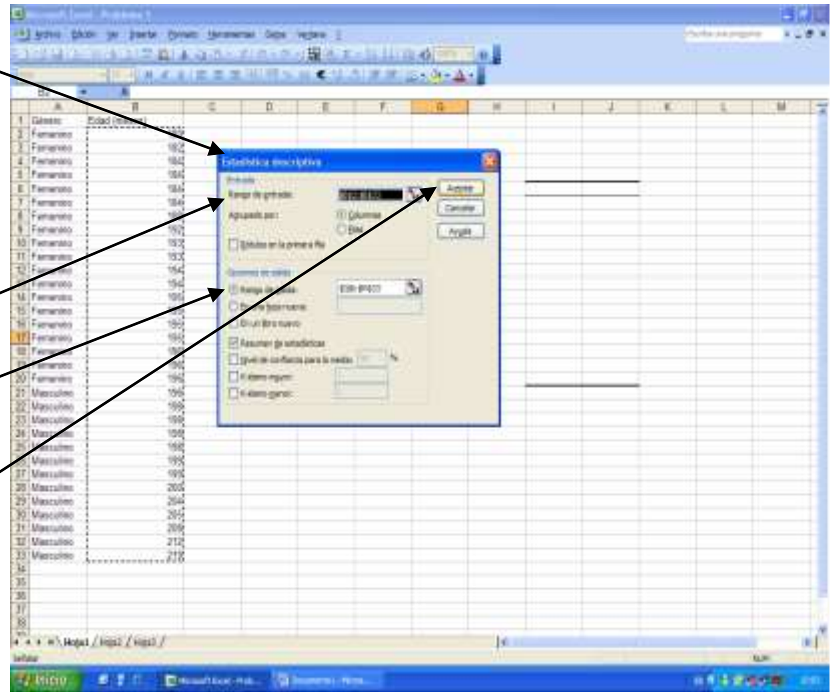


Ahora vamos a calcular lo estudiado en las clases anteriores:

Clic en:
El botón Herramientas en la Barra de Menú.
Análisis de datos.
En los comandos que se despliegan se Selecciona Estadística Descriptiva.
Y pulsar **Aceptar**



El **cuadro de diálogo** servirá para realizar un análisis descriptivo. En este caso consiste en especificar las celdas que contienen los números a los que se les quiere calcular las medidas estudiadas.

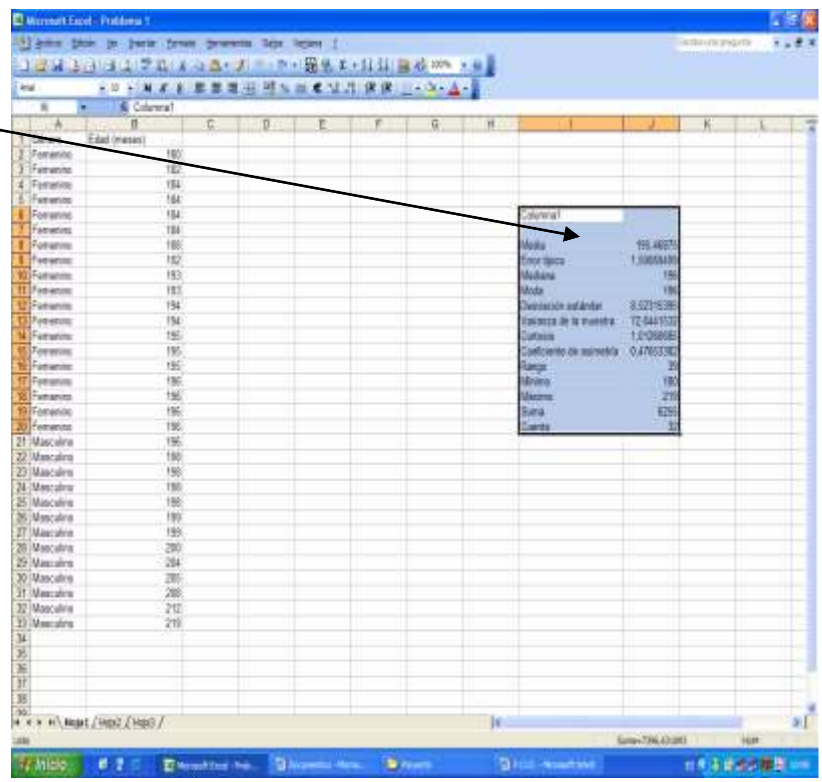


<u>Ejemplo:</u> para la variable edad (meses) en **Rango de entrada.** Seleccionar las celdas desde la B2 hasta la B33.

- En **Rango de salida** seleccionar en celdas vacías 16 filas y 2 columnas, en ese lugar se van a mostrar los resultados.
- Y pulsar **Aceptar**

El resultado se muestra como aparece en el cuadro en azul

Columna1	
Media	195,46875
Error típico	1,50669499
Mediana	196
Moda	196
Desviación estándar	8,52315395
Varianza de la muestra	72,6441532
Curtosis	1,01260685
Coeficiente de asimetría	0,47653382
Rango	39
Mínimo	180
Máximo	219
Suma	6255
Cuenta	32



Se ha calculado la media, mediana, moda, desviación estándar, varianza y rango, los otros resultados estadísticos no serán estudiados.

Problemas de consolidación

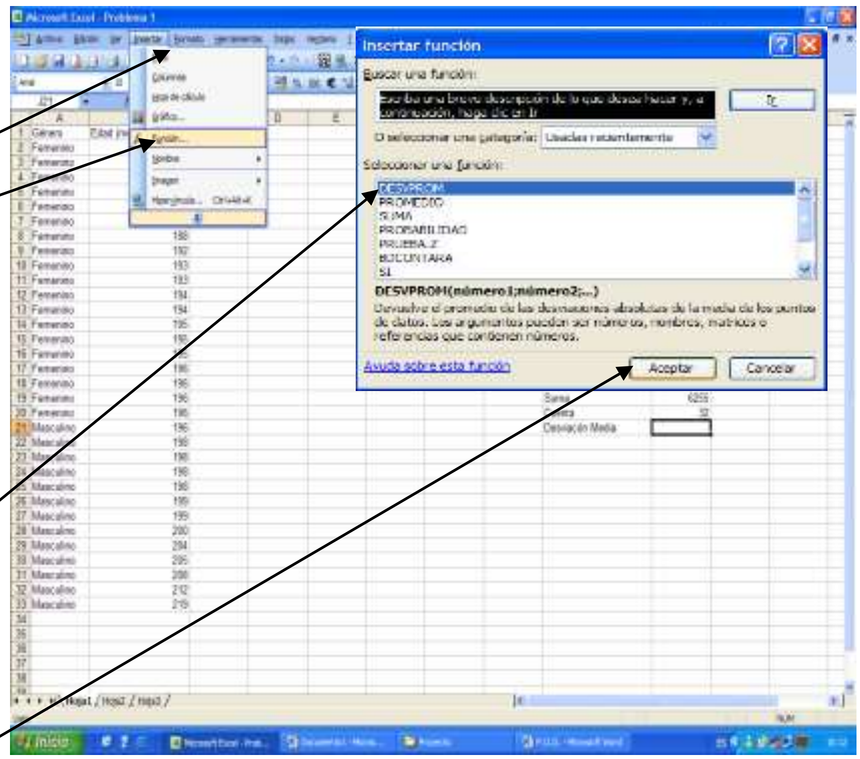
1. Interpretar las medidas que se han estudiado para los datos de la variable edad (meses).
2. Calcular e interpretar todas las medidas posibles para los datos de la variable promedio aritmético de notas (puntos) de la **TABLA A**.

Vamos a calcular la desviación media (DM).

El programa Excel cuenta con gran número de funciones predefinidas a las que se accede fácilmente haciendo clic en el botón **Insertar** de la Barra de Menú y clic en **función**.

Se sitúa en la celda donde queremos colocar la fórmula y donde se mostrará el resultado de la función, en nuestro ejemplo vamos a calcular la desviación media llamada en Excel **DESV. PROM.**

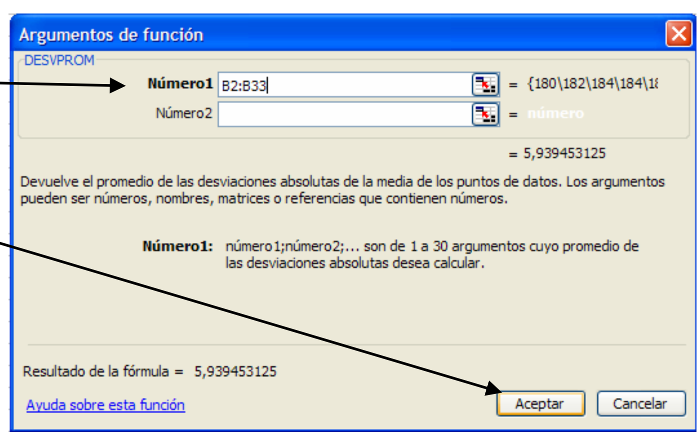
El resultado se quiere que aparezca en la celda J21 y en la celda I21 se escribe Desviación Media y queremos y se pulsa **Aceptar**



En **Número 1** se seleccionan los valores introducidos que van desde las celdas B2 hasta la B33.

Y pulsar **Aceptar.**

Se obtiene el resultado en la celda J21 antes señalada de este modo.

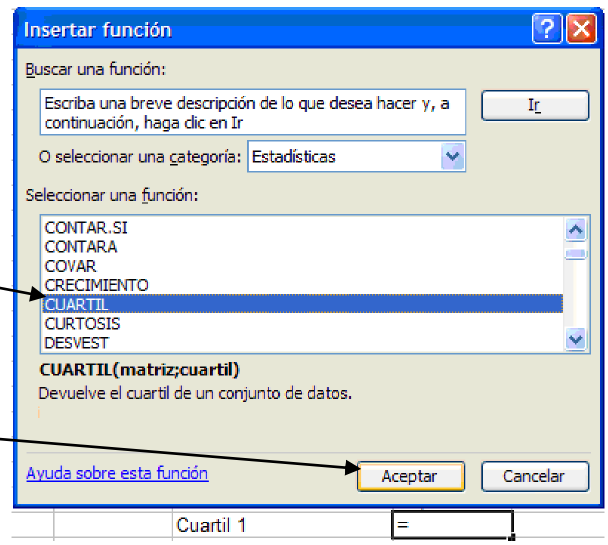


Problemas de consolidación

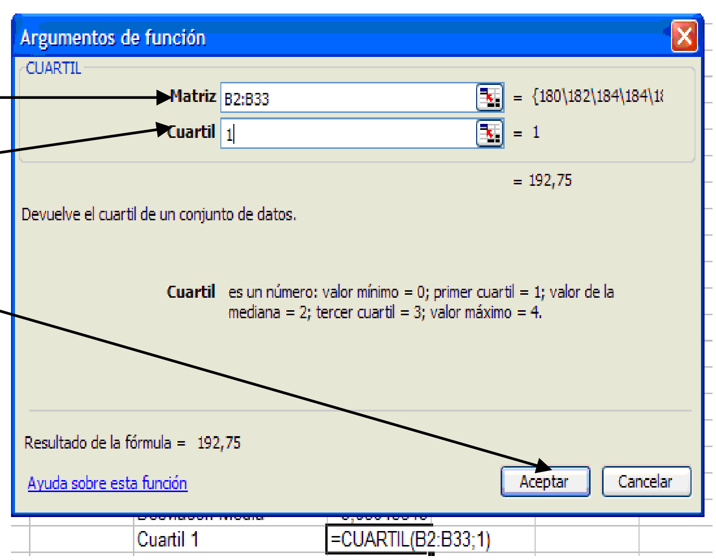
1. Interpretar la desviación media de los datos de la variable edad (meses) de la **TABLA A**.
2. Calcular e interpretar la desviación media de los datos de la variable promedio aritmético de notas (puntos) de la **TABLA A**.

Veamos cómo se calcula el cuartil 1 (Q_1).

- Clic en el botón Insertar de la Barra de Menú y luego en función.



- Situarse en la celda donde se quiere colocar el nombre del estadístico que se va a calcular y la celda donde se mostrará el resultado de la función, en este ejemplo se va a calcular el cuartil 1, por eso se utiliza la función **CUARTIL**.
- En la celda I22 escribimos Cuartil 1 y el resultado se quiere que aparezca en la celda J22.
- Y clic en **Aceptar.**



- En **Matriz** se seleccionan los valores introducidos que van desde las celdas B2 hasta la B33.
- En **Cuartil** el que se va a calcular, bien sea 1, 2 o 3.
- Y clic en **Aceptar.**
- Se obtiene el resultado en la celda J22 antes señalada de este modo.

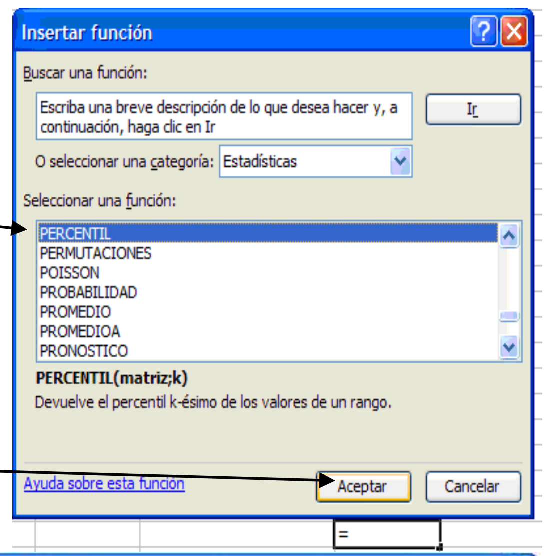
Problemas de consolidación

1. Interpretar el cuartil 1 de los datos de la variable edad (meses) de la **TABLA A**.
2. Calcular e interpretar el cuartil 3 para los datos de la variable edad (meses) de la **TABLA A**.
3. Con lo anterior calcular e interpretar el Rango Cuartil (RQ) de los datos de la variable edad (meses) de la **TABLA A**.
4. Calcular e interpretar el Rango cuartil (RQ) para los datos de la variable promedio aritmético de notas (puntos) de la **TABLA A**.

Veamos cómo calcular el percentil 90 (P_{90}).

- Clic en el botón Insertar en la Barra de Menú y se selecciona función.

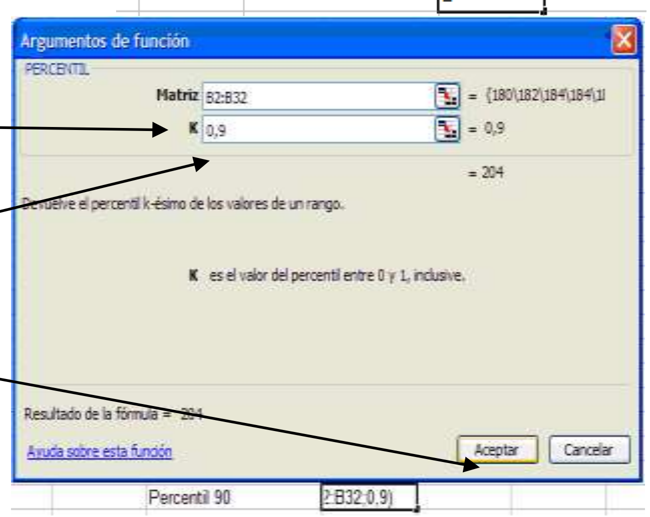
- Se sitúa en la celda donde se quiere colocar el nombre del estadístico que se va a calcular y la celda donde se mostrará el resultado de la función, en nuestro ejemplo vamos a calcular el percentil 90, por eso utilizamos la función **PERCENTIL.**



- El resultado se quiere que aparezca en la celda J23 y en la celda I23 se escribe Percentil 90.

- Y clic en **Aceptar.**

En **Matriz se** seleccionan los valores introducidos que van desde las celdas B2 hasta la B33.



- En **k** el percentil que se va a calcular, en el programa se toma del 0 al 1, por ello se anota 0,9.
- Y clic en **Aceptar.**
- Se obtiene el resultado en la celda J23 antes señalada de este modo.

Percentil 90	204,9

Problemas de consolidación

1. Interpretar el percentil 90 de los datos de la variable edad (meses) de la **TABLA A**.
2. Calcular e interpretar el percentil 10 para los datos de la variable edad (meses) de la **TABLA A**.
3. Con lo anterior calcular e interpretar el Rango Percentilico (RP) de los datos de la variable edad (meses) de la **TABLA A**.
4. Calcular e interpretar el Rango Percentilico (RP) para los datos de la variable promedio aritmético de notas (puntos) de la **TABLA A**.

5.3.2.- *Funcionamiento*

- Utilizar el esquema general de la Propuesta de Orientación Didáctica para ubicar el contenido a desarrollar.
- Desarrollar cada clase en dos horas académicas o noventa (90) minutos, porque la Ley Orgánica de Educación estipula que cada hora académica tiene una duración de cuarenta y cinco (45) minutos.
- Utilizar los problemas de consolidación como reforzamiento del aprendizaje adquirido para que continúe el proceso del desarrollo cognoscitivo, lógico – matemático en el aprendizaje obtenido (Londoño, 1995).
- Evaluar los contenidos con una prueba de desarrollo, porque permite observar aspectos conceptuales y procedimentales, dando paso a la comprensión e internalización de conocimientos estadísticos adecuados (Batanero, 2001).

5.3.3.- *Fases*

- Conocimiento de la Propuesta de Orientación Didáctica.
- Ejecución o puesta en marcha de la Propuesta de Orientación Didáctica.
- Evaluación del proceso de ejecución.
- Recopilación de datos obtenidos.
- Análisis de resultados.

- Conclusiones y recomendaciones.

5.4.- *Personal requerido*

- Coordinador(es) de la investigación.
- Docentes de Matemática.

5.5.- *Recursos necesarios para su puesta en práctica*

- Aulas de clase con su pizarrón.
- Aula de informática.
- Ejemplares de la propuesta.
- Marcadores acrílicos o tiza.
- Borradores de pizarra.

5.6.- *Estudios de costos y financiamiento*

Costos

Erogaciones en que se incurrirán para llevar a cabo la propuesta en las Unidades Educativas pilotos. Socialización y capacitación a los docentes: a través de conferencias se dará a conocer a las diferentes unidades educativas la propuesta, incluyendo al personal directivo, docentes y estudiantes de séptimo, octavo, noveno y segundo año. Se deben tener en cuenta espacios físicos, ayudas audiovisuales y materiales de apoyo. Módulos: se deben entregar impresos y en medios magnéticos los contenidos del programa a cada unidad educativa seleccionada previamente. Computadores o el servicio de computadores: es necesario para la última clase contar con un equipo de computación por estudiante, si no los tiene la unidad educativa se debe alquilar el servicio o buscar el organismo que pueda contribuir con el servicio (privadas o públicas).

Financiamiento

Los recursos para financiar los gastos se pueden solventar por diferentes fuentes:

- Unidad Educativa: verificar si cuenta en sus finanzas con los recursos monetarios para respaldar la propuesta.

- Estudiantes: confirmar si tienen capacidad económica para colaborar con algunos elementos utilizados en las diferentes actividades de acuerdo al contenido del módulo.
- Docentes: comprobar si tienen disposición para asumir ciertos gastos que se generan en la implantación del programa.
- Tesistas o investigadores: manifestar si los investigadores tienen la disponibilidad de financiar personalmente la puesta en marcha del programa en las unidades educativas seleccionadas.
- Entidades privadas: indagar si existen entes privados que apoyan actividades educativas tendientes a mejorar el nivel académico de la escuela aportando recursos económicos y/o físicos.
- Entidades públicas: constatar si los organismos públicos del sector educativo, tienen disponibilidad presupuestal para contribuir al desarrollo de programas en las unidades educativas tendientes a facilitar el proceso de la formación integral de los educandos, así como la posibilidad de asignar docentes con el perfil requerido en caso de que la unidad educativa lo requiera.

CAPÍTULO 6. Discusión de resultados

6.1.- *Factibilidad del modelo propuesto*

Los cinco (5) docentes de Matemática que validaron el módulo consideraron que el 100% estaba bueno en cuanto a los seis criterios de evaluación. Manifestaron de manera oral que es factible este modelo propuesto debido a su sencillez lo cual hace que cualquier docente tenga la capacidad de utilizarlo sin encontrar mayor dificultad. La Estadística debe ser enseñada con este módulo por ser muy interesante, dinámico y contextualizado, lo cual les proporcionará un buen aprendizaje a los estudiantes en esta ciencia. Se debe enseñar al comienzo del año escolar para darle mayor importancia y porque esta presentada de manera muy motivadora. La última clase de la propuesta muestra que la Estadística no está aislada de los ordenadores demostrando la valía de este módulo. Uno de los evaluadores mencionó lo largo de la clase número siete, sin embargo, aseguró que lo atrayente de la propuesta es que no se necesitan recursos fuera de lo común y lo más difícil sería la aplicación de la última clase porque no todas las instituciones tienen un aula de informática.

6.2.- *Control y evaluación de procesos*

Control

Deben evaluarse los contenidos con una prueba de desarrollo, ya que permite observar aspectos conceptuales y procedimentales y el (o los) coordinador(es) debe(n) revisar la evaluación de los profesores en períodos no mayores a quince (15) días, concluyendo con un registro general de todas las evaluaciones efectuadas.

Evaluación de procesos

Se debe evaluar por parte del profesor la aplicación de la propuesta a través de un instrumento medido en escala tipo Lickert y el (o los) coordinador(es) deben elaborar un instrumento que le permita resumir y luego analizar los resultados de todos y cada uno de los profesores en cuanto a la evaluación de los alumnos y la aplicación de la propuesta.

6.3.- Conclusiones, limitaciones y recomendaciones finales

Conclusiones

En este trabajo de investigación se determinó que la Estadística no es enseñada porque está al final del programa de estudio y los docentes no tienen el suficiente tiempo para incluirla en sus planes de lapso, lo cual despoja a los estudiantes del aprendizaje de esta ciencia, que es tan importante para interpretar y comprender situaciones problemas de la vida real.

El bajo nivel de motivación de los estudiantes, está por lo general asociado a la omisión de la enseñanza de la Estadística por parte de los docentes, debido a que estos consideran más importantes otros contenidos.

Es primordial conocer las fallas, deficiencias y fortalezas que presenta el currículo en los contenidos de Estadística de la Tercera Etapa de Educación Básica y del segundo año de Educación Media Diversificada, pues esto ayuda a que instituciones, docentes y analistas puedan buscar las soluciones adecuadas para solventar dicho problema y mejorar la práctica docente.

El Módulo para la enseñanza-aprendizaje de la Estadística en el bachillerato a través de situaciones problemas es factible que se aplique de acuerdo a los resultados de la validación.

Limitaciones

Es una propuesta muy extensa porque está planteada para los diferentes años escolares en que se debe impartir la Estadística según el currículo vigente, esto hace que por razones de tiempo sea difícil su aplicación y por tanto su evaluación para luego determinar su eficiencia.

Existen instituciones que no poseen todos los recursos necesarios para la puesta en marcha de la propuesta, además hay limitaciones en el recurso humano, por ello se debe buscar la colaboración de diversos organismos en su financiamiento.

Recomendaciones finales

Las instituciones, docentes y alumnos deben reconocer la importancia que reviste la enseñanza-aprendizaje de la Estadística en el bachillerato.

Los contenidos de la Estadística deben ser desarrollados de manera contextualizada considerando el entorno de la institución, la realidad socioeconómica del país y la posibilidad de que los alumnos participen de manera activa en su aprendizaje para lograr solventar los problemas pedagógicos presentados para su enseñanza - aprendizaje.

Utilizar el Módulo para la enseñanza de los contenidos estadísticos con el fin de promover un aprendizaje significativo en los estudiantes que transitan por esta etapa del bachillerato.

Emplear el Módulo para futuras investigaciones que permitan llegar a conocer la eficiencia o eficacia del modelo propuesto.

REFERENCIAS BIBLIO - HEMEROGRÁFICAS

Armas, J. (1988). Estadística Sencilla: Descriptiva. Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Batanero, C. (2001). Didáctica de la Estadística. Trabajo de investigación del Departamento de Didáctica de la Matemática .Universidad de Granada.

Batanero, C. y Godino, J. (2001). Análisis de datos y su didáctica. Trabajo de investigación del Departamento de Didáctica de la Matemática .Universidad de Granada.

Barrera, C. (2004). Planificación Educativa. Extraído el 16 de Junio de 2008 de http://www.geocities.com/umsada/trabajoar8.htm.

Constitución Nacional de la República Bolivariana de Venezuela (1999). Gaceta oficial de la República de Venezuela, Nº 36860 (Extraordinaria).

Behar, R. y Grima P. (2001). Mil y una dimensiones del aprendizaje de la Estadística. España: Estadística española.

Da Silva, (S.F.). Estadística. Extraído el 5 de enero del 2008 de http://www.monografías.com/trabajos10/esta/esta.shtml.

Devia, R. y Mora M. (2007). Propuesta para la enseñanza de las nociones elementales de estadística descriptiva mediante el uso del software SPSS como recurso de orientación didáctica en alumnos de 9º grado de E.B. Memoria de Grado presentada para optar al Titulo de Licenciado en Educación Matemáticas, Facultad de Humanidades y Educación de la Universidad de Los Andes, Mérida, Venezuela.

Medina, D. (2006). Desarrollo de un software educativo para la enseñanza de la Estadística en el séptimo grado de educación básica. Trabajo de grado de maestría publicado. Universidad de Los Andes. Mérida. Venezuela.

Levin y otros (2004). Estadística para Administración y Economía. México: Pearson Educación.

Londoño, P. (1995). El lenguaje, la matemática y la enseñanza de las operaciones básicas. Trabajo de Ascenso publicado. Universidad de Los Andes. Mérida. Venezuela.

Hernández, R., Fernández, C., y Baptista, P. (2003). Metodología de la Investigación. México D.F: Mc Graw Hill.

Moreno, A. y Vallecillos, A. (2002). La inferencia estadística básica en la enseñanza secundaria. Trabajo de investigación de la Universidad de Granada. España. Extraído el 15 de junio del 2008 de http://www.caib.es/ibae/esdeveniment/jornades_10_01/doc/jornadas%20europeas%20de%estadistica2001.doc.

Hurtado, J. (2000). Metodología de la Investigación Holística. Caracas: Fundación Sypal.

Parra, H. (1995). Resolución de problemas, eje central de la enseñanza de las matemáticas. Caracas: Movimiento Pedagógico.

Piaget, J. y Beth, E. (1980).Epistemología, Matemática y Psicología. Barcelona: Grijalbo.

Pestaña, P. (2002). Conceptos básicos, terminología y metodología de la estadística. Caracas: CAC, S.A.

Sabino, C. (1992). El proceso de investigación. Caracas: Panapo.

Sin autor. (2002). Convocatoria SAGARPA. Extraído el 15 de junio del 2008 de http://planeación.cicese.mx/convocatorias/2002-SAGARPA/referencia/htm).

Torres, E. (1997). Nociones de Estadística Descriptiva. Trabajo de Ascenso a Profesor Asociado. Universidad de Los Andes. Mérida. Venezuela.

Anexo A. *Formato de la entrevista*

Universidad de Los Andes
Facultad de Humanidades y Educación
Escuela de Educación
Mención Matemática

Guía de entrevista

Fecha: _____/_______/______

Datos del encuestado:

Especialidad: ________________________Años de experiencia: ___________________

Estimado encuestado: Todas las preguntas o ítems de esta entrevista, están orientados a obtener información sobre la enseñanza – aprendizaje de la Estadística que han impartido en sus años anteriores de experiencia, por favor responda de acuerdo a su criterio lo que considere pertinente.

Preguntas o ítems:

1. En su labor como docente de la Matemática ¿Usted ha enseñado Estadística?

Si____ No____

¿Porque?____
____
____
____
____
____

2. ¿Qué dificultades más comunes encontró en la en la enseñanza de la Estadística en el bachillerato? Especifique

____
____
____
____
____

3. ¿Considera qué es importante la enseñanza de los contenidos de Estadística en el bachillerato?

Si____ No____

¿Porque?____

____

____

____

____

____

____

____

4. ¿Es conveniente las estrategias de enseñanza de la Estadística sugerida en los libros de textos?

Si____ No____

¿Porque?____

____

____

____

____

____

____

____

Anexo B. *Prueba diagnóstica de conocimiento*

República Bolivariana de Venezuela
Ministerio del Poder Popular para la Educación
Área de Matemática
Colegio Salesiano "San Luis"

Prueba diagnóstica de conocimiento

Datos del alumno:
Grado: ___________Sección: _______ Fecha: __________________

Contenidos a evaluar: conceptos básicos y medidas de tendencia central de Estadística.

Estimado alumno:
A continuación se presentan las siguientes instrucciones generales para responder la prueba:
1). Lea con detenimiento las instrucciones de cada una de las partes de la prueba y sus respectivos ítems antes de responderlos.
2). En caso de duda, consulte a la persona que aplica la prueba.
3). Trate de responder la prueba en su totalidad.
4). Usted dispone de treinta minutos para responder la totalidad de la prueba.

Parte 1. *Las siguientes son preguntas de selección múltiple con una única respuesta correcta, encierre la respuesta que considere correcta. (Valor: 2 puntos c/u)*

1. La ciencia que estudia las características de un conjunto de casos para hallar en ellos regularidades en el comportamiento que sirvan para describir el conjunto y para efectuar predicciones es la:

a) Población.
b) Muestra.
c) Estadística.
d) Categoría.

2. Cada uno de los grupos básicos en los que puede incluirse o clasificarse los valores que asume una variable estadística se denominan:

a) Categorías.
b) Muestras.
c) Datos.
d) Variables estadísticas.

3. El conjunto de datos o elementos que cumplen ciertas propiedades comunes es:

a) Una categoría.
b) La población.
c) Un dato.
d) Una variable estadística.

4. El peso en Kilogramos de los estudiantes del Colegio Salesiano "San Luis" es una:
a) Variable cualitativa.
b) Variable cuasi – cuantitativa.
c) Variable cuantitativa discreta.
d) Variable cuantitativa continua.

5. El nivel socio - económico de los estudiantes del Colegio Salesiano "San Luis" es una:
a) Variables cualitativas.
b) Variables cuasi – cuantitativas.
c) Variables cuantitativas discretas.
d) Variables cuantitativas continuas.

6. El número de veces que se repite un dato lleva por nombre:
a) Frecuencia absoluta.
b) Frecuencia acumulada.
c) Frecuencia relativa simple.
d) Frecuencia relativa acumulada.

7. El sexo de los estudiantes del Colegio Salesiano "San Luis" es una:
a) Variable cualitativa.
b) Variable cuasi – cuantitativa.
c) Variable cuantitativa discreta.
d) Variable cuantitativa continua.

8. La proporción obtenida del número de datos que se repiten entre el total de datos de la muestra es una:
a) Frecuencia absoluta.
b) Frecuencia acumulada.
c) Frecuencia relativa simple.
d) Una frecuencia relativa acumulada.

9. La moda de los puntajes: 17, 15, 15, 16, 14 es:
a) 14
b) 16
c) 17
d) 15

10. La media aritmética de los puntajes: 17, 15, 16, 18, 14 es:
a) 15
b) 17
c) 16
d) 14

Anexo C. *Instrumento para determinar la validez de la prueba diagnóstica*

Universidad de Los Andes
Facultad de Humanidades y Educación
Escuela de Educación
Mención Matemática

Instrumento para determinar la validez de la prueba de conocimiento

Objetivo general

El objetivo del presente instrumento es determinar la validez de la prueba de conocimiento aplicada a estudiantes del segundo año del Ciclo de Educación Media Diversificada, utilizando el método de Juicios de Expertos, el cual permitirá establecer la validez de contenido de cada ítem, la validez de contenido de todo el instrumento y el nivel de concordancia entre los jueces, a través del Coeficiente de Proporción de Rangos (CPR); para tal efecto se necesita de su colaboración en el proceso de evaluación de cada uno de los ítems del instrumento.

Instrucciones

Anote sus datos, incluyendo nombres y apellidos, profesión y años de experiencia.

El proceso de calificación de los ítems se realiza por medio de una escala del uno al tres, donde el 1=deficiente; 2=regular y 3=bueno, marque con una equis (X) la calificación correspondiente a cada uno de los ítems presentados en la prueba de conocimiento.

Se presenta una tabla que contiene cinco columnas, donde:

- Los números en la primera columna se refieren a los ítems de la prueba de conocimiento, el cual se anexa al presente instrumento.

- La segunda columna, nombrada 1 corresponde a los ítems de la prueba que obtengan una calificación deficiente, porque no existe claridad en la redacción, pertinencia de los ítems, estructura gramatical y consistencia de las alternativas.

- La tercera columna, designada 2 corresponde a los ítems de la prueba que obtengan una calificación regular, porque falta claridad en la redacción, pertinencia de los ítems, estructura gramatical y consistencia de las alternativas.

- La cuarta columna, denominada 3 corresponde a los ítems de la prueba que obtengan una calificación buena, porque existe claridad en la redacción, pertinencia de los ítems, estructura gramatical y consistencia de las alternativas.

- La quinta columna, titulada observaciones se relaciona con las sugerencias de cada ítem.

Criterios de validación:

- Pertinencia de los ítems, referida concretamente a que los reactivos se ajusten al contenido a evaluar.

- Claridad en la redacción, no debe lugar a reactivos que se presten para una interpretación ambigua.

- Estructura gramatical, lo suficiente clara y precisa y ajustado al nivel académico de los estudiantes que cursan el segundo año del Ciclo de Educación Media Diversificada.

- Consistencia de las alternativas, las diferentes alternativas deben ser pertinentes respecto a cada ítem.

Validación de la prueba de conocimiento

Juez:

Nombres y apellidos: ____

Profesión: ____

Años de experiencia: ____

Ítems	*Calificación*			*Observaciones*
	1	*2*	*3*	
Ítem 1				
Ítem 2				
Ítem 3				
Ítem 4				
Ítem 5				
Ítem 6				
Ítem 7				
Ítem 8				
Ítem 9				
Ítem 10				

____

Firma

HUMANIDADES Y EDUCACIÓN
UNIVERSIDAD DE LOS ANDES
MÉRIDA – VENEZUELA

Constancia de validación

Yo, ____C.I. _______________________ de profesión: ____, en mi condición de la presente hago constar que he revisado, con fines de validación, la prueba diagnóstica de conocimiento para ser aplicada a estudiantes el segundo año del Ciclo de Educación Media Diversificada elaborada por los bachilleres Joan F. Chipia L. y Carmen Z. Lara A.

____

Firma

Anexo D. *Validez de la prueba diagnóstica*

El procedimiento estadístico empleado para determinar la validez de contenido del instrumento es el denominado Coeficiente de Proporción de Rangos (CPR). Este se determina usando el método de juicio de expertos, obteniéndose con este coeficiente que permite determinar la validez de contenido de cada ítem, la validez de contenido de todo el instrumento y el nivel de concordancia entre los jueces (Hernández, 1995).

Los resultados arrojados son:

El promedio de rangos

$$PRi = \frac{\sum ri}{j} = \frac{88}{3} = 29{,}33$$

La relación proporcional

$$CPRi = \frac{PRi}{vmr} = \frac{29{,}33}{3} = 9{,}776$$

La probabilidad del error

$$Pe = \left[\left(\frac{1}{j}\right)\right]^{j} = \left[\left(\frac{1}{3}\right)\right]^{3} = 0{,}036$$

El Coeficiente de Proporción de Rangos Corregido por Concordancia:

$$CPRic = CPRi - Pe = 9{,}776 - 0{,}036 = 9{,}74$$

El Coeficiente de Proporción de Rangos Corregido por Concordancia aleatoria:

$$CPRt = \frac{\sum CPRic}{N} = \frac{9{,}78}{10} = 0{,}978$$

$$CPRtc = CPRt - Pe = 0{,}978 - 0{,}036 = 0{,}942$$

Estimación del Coeficiente de Proporción de Rangos de la
prueba diagnóstica de conocimiento.

Nº de ítem	*Juez 1*	*Juez 2*	*Juez 3*	$\sum ri$	PRi	$CPRi$
1	3	2	3	8	8 /3 = 2,67	2,67 / 3 = 0,89
2	3	3	3	9	9 /3 = 3	3 / 3 = 1
3	3	3	2	8	8 /3 = 3	2,67 / 3 = 0,89
4	3	3	3	9	9/3 = 3	3 / 3 = 1
5	3	3	3	9	9 /3 = 3	3 / 3 = 1
6	3	3	3	9	9 /3 = 3	3 / 3 = 1
7	3	3	3	9	9 /3 = 3	3 / 3 = 1
8	3	3	3	9	9 /3 = 3	3 / 3 = 1
9	3	3	3	9	9 /3 = 3	3 / 3 = 1
10	3	3	3	9	9 /3 = 3	3 / 3 = 1

Fuente: Joan F. Chipia L. y Carmen Z. Lara A.

El cálculo del CPR arroja que la validez y concordancia es excelente, por encontrarse entre 0,90 y 1,00.

Anexo E. *Instrumento de validación del Módulo*

Universidad de Los Andes
Facultad de Humanidades y Educación
Escuela de Educación
Mención Matemática

Instrumento de validación del Módulo para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas

Objetivo general

El objetivo del presente instrumento es validar el diseño del Módulo a través de la Propuesta de Orientación Didáctica para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas (Municipio Libertador-Mérida), para tal efecto se necesita de su colaboración en el proceso de validación de cada uno de los criterios por clase presentados en la tabla 1.

Instrucciones

1. A continuación se presenta la tabla 1 de doble entrada que está estructurada por seis columnas y doce filas, donde:

- Las filas corresponden a las doce clases presentadas en el Módulo, la cual se anexa al presente instrumento.
- Las columnas hacen referencia a los criterios para validar el Módulo.

2. Valide el Módulo marcando con una equis (X) la calificación de cada clase, considerando cada uno de los criterios presentados en la tabla 1, donde: D = deficiente, R = regular, B= bueno.

NOTA: Si la calificación obtenida es deficiente o regular se debe realizar la observación correspondiente en la tabla 2.

3. Anote sus datos, incluyendo nombres y apellidos, profesión y años de experiencia.

<u>Criterios de validación:</u>

- Presentación, constituye el cuerpo o forma del Módulo.
- Secuencia conceptual, referida al encadenamiento y organización pertinente de los conceptos a desarrollar.
- Ejemplos ilustrados, referida a la utilización de dibujos y/o diagramas para modelar un tópico.
- Problemas contextualizados, se refiere a situaciones que involucren el medio ambiente en el que se desenvuelven los estudiantes.
- Estrategia didáctica, correspondiente al método de enseñanza empleado.
- Fundamento matemático, referido al basamento que nos permite construir los contenidos de la Estadística.

Tabla 1. Criterios de evaluación por clase

Criterios / Nº de Clases	Presentación			Secuencia conceptual			Ejemplos ilustrados			Problemas contextualizados			Estrategia didáctica			Fundamento matemático		
	D	R	B	D	R	B	D	R	B	D	R	B	D	R	B	D	R	B
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		

Nombres y Apellidos ____

Especialidad: _______________________________Años de experiencia: _________

____

Firma del Evaluador

HUMANIDADES Y EDUCACIÓN
UNIVERSIDAD DE LOS ANDES
MÉRIDA – VENEZUELA

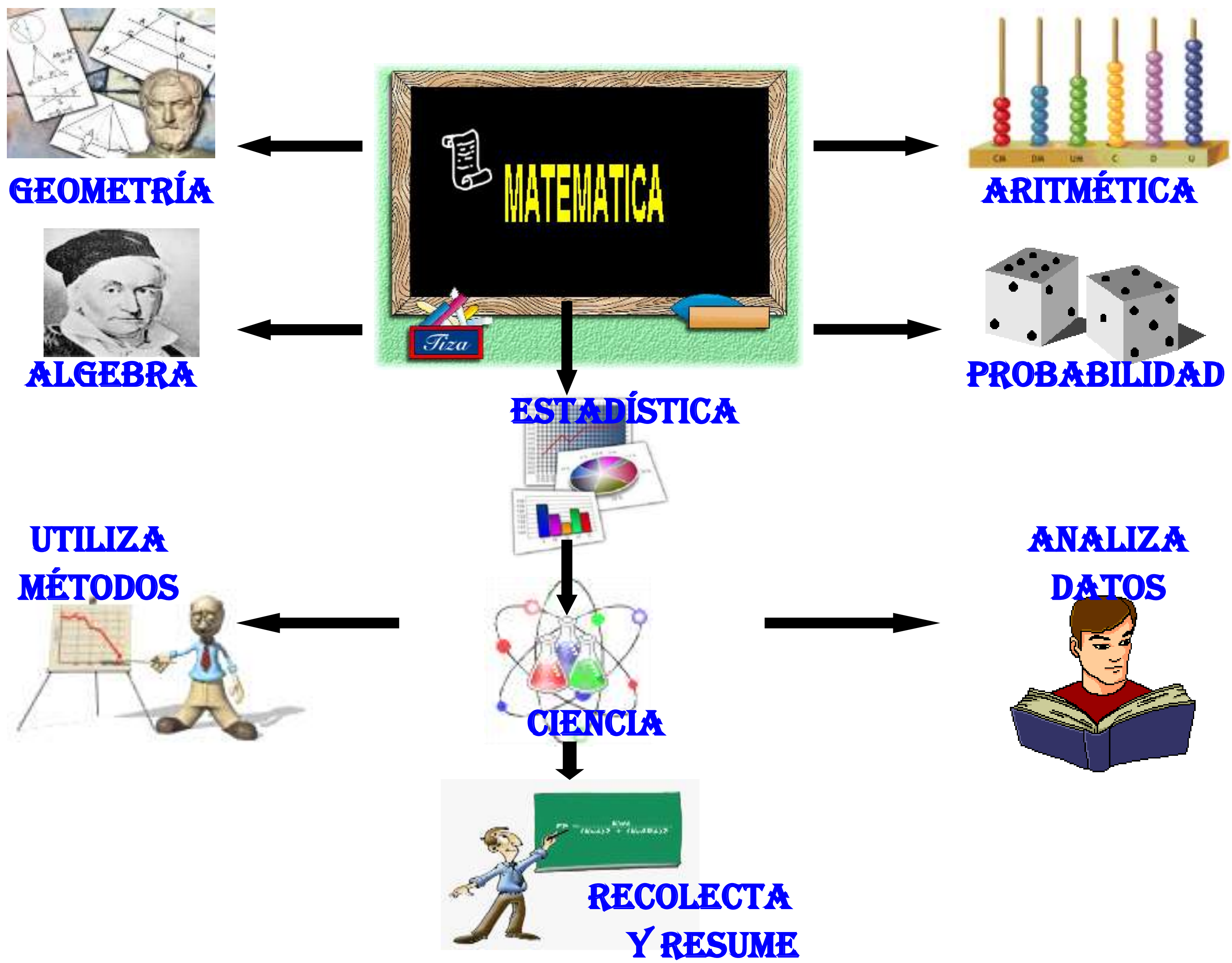
Constancia de validación

Yo, ____C.I. ______________________ de profesión: ____, en mi condición de la presente hago constar que he revisado y validado por medio del instrumento presentado el Módulo a través de una Propuesta de Orientación Didáctica para la enseñanza – aprendizaje de la Estadística en el bachillerato a través de situaciones problemas elaborada por los bachilleres Joan F. Chipia L. y Carmen Z. Lara A.

____

Firma

Anexo F. *Mapa conceptual general*



Anexo G. *Mapa conceptual específico*