

DISTRIBUCIÓN DE FRECUENCIAS E HISTOGRAMAS

GUÍA DEL MAESTRO

Autor: Dr. Edwin Morera González

Materia: Matemáticas **Nivel:** 7-9.

Conceptos Principales: Distribución de frecuencias, histogramas.

Conceptos Secundarios: Población, muestra, datos y tipos de datos.

Conocimiento previo: Conjuntos, subconjuntos, promedios, gráficas de barras.

Objetivos específicos:

Durante la actividad el participante:

1. Define los siguientes conceptos:
 - a. Nivel de medición nominal
 - b. Nivel de medición ordinal
 - c. Nivel de medición de intervalo
 - d. Nivel de medición de razón
 - e. Muestra
 - f. Población
 - g. Muestra aleatoria
 - h. Distribución de frecuencias
 - i. Límites de clase inferiores
 - j. Límites de clase superiores
 - k. Fronteras de clase
 - l. Marcas de clase
 - m. Anchura de clase
2. Determina el nivel de medición de una colección de datos.
3. Identifica la población en una situación dada.
4. Identifica el tipo de muestra en una situación dada.
5. Identifica la anchura de clase, las marcas de clase y las fronteras de clase para una distribución de frecuencias dado.
6. Elabora la distribución de frecuencias de un conjunto de datos.
7. Elabora la distribución de frecuencias relativas de un conjunto de datos.

Estándares, Expectativas e Indicadores por Grado:

ESTÁNDAR DE CONTENIDO 5: ANÁLISIS DE DATOS Y PROBABILIDAD

El estudiante es capaz de utilizar diferentes métodos de recopilar, organizar, interpretar y presentar datos para hacer inferencias y conclusiones.



Séptimo

16.0 Formula preguntas sobre poblaciones pequeñas que pueden contestarse por medio de la recolección y análisis de datos de dos variables, diseños relacionados con investigaciones de datos y la recolección de datos.

E.RD.7.16.1 Formula una pregunta simple que involucre dos atributos.

E.RD.7.16.2 Define una pequeña población donde los datos pueden ser recolectados para contestar una pregunta.

E.RD.7.16.3 Identifica, selecciona, crea y utiliza varias formas de representar conjuntos de datos.

E.RD.7.16.4 Identifica dos atributos donde recolectar los datos, decide cómo medir estos atributos para responder la pregunta formulada y determina el proceso de recolección de datos.

17.0 Organiza y resume datos de dos variables, examina los datos de estos atributos y clasifica cada atributo como variable categórica o variable numérica.

E.RD.7.17.1 Clasifica cada atributo como variable cuantitativa o cualitativa

E.AD.7.17.2 Describe la distribución de cada atributo separadamente utilizando las gráficas apropiadas, (incluyendo diagramas de tallo y hoja, histogramas, diagramas de caja y resumen estadístico, incluyendo rango intercuartil.

E.RD.7.17.3 Identifica, describe y construye gráficas para representar datos de dos variables (tablas para dos variables, diagramas de caja paralela, diagramas de tallo y hoja dobles para una variable categórica y una variable numérica; y diagramas de dispersión, con la línea de tendencia apropiada.

E.AD.7.17.4 Explica las ventajas de las diferentes formas de representar datos.

E.AD.7.17.5 Describe la relación entre dos variables y los efectos de los extremos en las relaciones observadas.

18.0 Interpreta los resultados y comunica las conclusiones de los análisis de datos de dos variables para contestar la pregunta formulada utilizando los símbolos, notación y terminología apropiada.

E.AD.7.18.1 Interpreta y comunica las conclusiones de un análisis estadístico en dos variables en el contexto de la pregunta formulada utilizando la terminología apropiada.

E.AD.7.18.2 Identifica gráficas engañosas

19.0 Determina el espacio muestral para un experimento y determina, cuando sea posible, la probabilidad teórica para un evento definido en el espacio muestral. Describe y aplica la Regla de la Suma de probabilidades.

E.PR.7.19.1 Determina el espacio muestral para un experimento y utiliza listas, tablas y diagramas de árbol para representar los resultados posibles.

E.PR.7.19.2 Identifica los eventos para un espacio muestral dado, representa relaciones entre los eventos usando diagramas de Venn y determina las probabilidades para eventos y sus complementos.

E.PR.7.19.3 Describe y aplica la Regla de la Suma de probabilidades para eventos que son mutuamente exclusivos y eventos que no.

Octavo

13.0 Formula preguntas que pueden atenderse a través de la recolección y análisis de datos obtenidos de una encuesta. Evalúa los resultados de una encuesta presentada en los medios de comunicación.

E.RD.8.13.1 Formula una pregunta de interés y define los componentes claves que pueden atenderse a través de una encuesta.

- E.RD.8.13.2** Define la población, las variables que se medirán, y cómo se medirán e identifica los factores que pueden influir en los resultados de la encuesta.
- E.RD.8.13.3** Diseña cuestionarios.
- E.AD.8.13.4** Describe las técnicas para obtener muestras aleatorias simples de los miembros de una población.
- E.PR.8.13.5** Identifica situaciones donde un muestreo aleatorio estratificado de una población sería preferible a un muestreo aleatorio simple.
- E.PR.8.13.6** Identifica y describe las diferencias entre una muestra y un censo, y explica las ventajas y desventajas de cada uno.
- E.PR.8.13.7** Diseña e implementa la selección de una muestra aleatoria simple de una población, recolecta y organiza los datos; representa los datos en tablas y gráficas y resume los datos por medio de medidas de tendencia central y dispersión (incluyendo desviación absoluta media).
- E.RD.8.13.8** Describe como el método de seleccionar los sujetos para una muestra y los métodos de medición de los resultados afectan los resultados de la encuesta. Explica como pueden surgir sesgos de los errores de muestreo y errores de medición.
- E.AD.8.13.9** Examina los resultados de las encuestas presentadas en los medios de comunicación, discutiendo y evaluando cómo la muestra fue seleccionada de la población y los métodos utilizados para medirla, recolectarla y representarla. Identifica las fuentes de sesgos que pueden afectar los resultados de la encuesta.

14.0 Analiza, resume y compara los resultados de muestras aleatorias y no aleatorias y del censo, usando resúmenes estadísticos y una variedad de representaciones gráficas para comunicar sus hallazgos.

- E.AD.8.14.1** Compara las medidas de tendencia central y dispersión obtenidos de los datos de la muestra de una población (estadística) con las medidas de centro y dispersión obtenidos de los datos de un censo de la población (parámetros). Observa que los medios de la muestra tienden a acercarse a la media de la población a medida que le tamaño de la muestra aumente.
- E.AD.8.14.2** Reconoce que las medidas de tendencia central y dispersión obtenidas de muestras aleatorias pueden diferir de muestra a muestra aún si se obtienen de la misma población y tienen el mismo número de observaciones.
- E.AD.8.14.3** Distingue entre métodos de muestreo aleatorio y no aleatorio. Compara los resultados de muestras aleatorias y no aleatorias simples de la misma población; discute cómo y por qué los resultados pueden diferir debido a fuentes potenciales de sesgos en las muestras.

Noveno

9.0 Determina el espacio muestral de un experimento, y emplea la regla de conteo de multiplicación. (Propiedad Fundamental de Conteo).

- E.PR.9.9.1** Utiliza listas, tablas y diagramas de árbol para representar los resultados posibles en un espacio muestral para un experimento.
- E.PR.9.9.2** Emplea estrategias sistemáticas de conteo, como la Propiedad Fundamental de Conteo, para determinar el número de resultados posibles.
- E.PR.9.9.3** Distingue entre sucesos compuestos dependientes e independientes y explica la idea de probabilidad condicional.

E.PR.9.9.4 Diseña y utiliza árboles, tablas, modelos de área y otras representaciones para calcular la probabilidad de sucesos compuestos cuando los sucesos son independientes y cuando no lo son.

E.PR.9.9.5 Describe y aplica la regla de multiplicación para probabilidad para calcular probabilidades para sucesos compuestos dependientes y para independientes.

10.0 Desarrolla, usa e interpreta simulaciones para estimar probabilidades para eventos cuyos valores teóricos son difíciles o imposibles de calcular.

E.PR.9.10.1 Describe una simulación identificando los componentes y supuestos en un problema, seleccionando un instrumento para generar los resultados, define intento, y especifica el número de intentos; y conduce la simulación.

E.PR.9.10.2 Resume datos de una simulación usando los resúmenes numéricos y las gráficas apropiadas, desarrolla un estimado para la probabilidad de un evento asociado a una situación probabilística del mundo real, y discute el efecto de un número de intentos en la probabilidad estimada de un evento.

E.PR.9.10.3 Reconoce que los resultados de una simulación difieren de una simulación a otra; observa que los resultados de una simulación tienden a converger a medida que aumenta el número de intentos.

11.0 Analiza datos numéricos en dos variables, representando estos datos con diagramas de dispersión apropiadas y traza la línea de mejor ajuste.

E.AD.9.11.1 Juzga si el diagrama de dispersión aparenta demostrar tendencias lineales, y si es así, traza la línea de mejor ajuste y escribe la ecuación de esta línea; usa la ecuación para establecer predicciones; e interpreta la pendiente de la línea en el contexto del problema.

E.AD.9.11.2 Calcula la línea de mejor ajuste, a mano para modelar una relación representada en un diagrama de dispersión, e interpreta la pendiente e intercepto en términos del contexto del problema.

Trasfondo: ¿Qué es la estadística?

La estadística es el lenguaje universal de la ciencia. Como usuarios potenciales de la estadística, es necesario dominar la “ciencia” y el “arte” de utilizar correctamente su metodología. El empleo cuidadoso de los métodos estadísticos permite obtener información precisa de los datos. Estos métodos incluyen: 1. definir cuidadosamente la situación, 2. recolectar datos, 3. resumir con precisión los datos y 4. obtener y comunicar las conclusiones importantes.

La estadística implica información, números y gráficas para resumir esta información, y su interpretación. El término *estadística* posee varios significados para personas de diversos entornos e intereses. Para algunos, es un campo de “magia” en que una persona con conocimientos supera a las demás. Para otros, se trata de un medio para recolectar y representar grandes cantidades de información. Y algunos entienden, que se trata de un medio para “tomar decisiones de frente a la incertidumbre”. En la perspectiva idónea, cada uno de estos puntos de vista es correcto.

El campo de estudio de la estadística puede dividirse a grandes rasgos en dos áreas: *estadística descriptiva* y *estadística inferencial*. La **estadística descriptiva** es en lo que la mayoría de las personas piensan al escuchar la palabra estadística. La estadística descriptiva incluye la recolección, presentación y descripción de datos muestrales. En otras palabras, es la parte de la estadística que sólo se ocupa de describir y analizar un grupo dado, sin sacar conclusiones sobre un grupo mayor. El término **estadística inferencial** se refiere a la técnica de interpretación de los



valores resultantes de las técnicas descriptivas y a la toma de decisiones y obtención de conclusiones sobre la población. Si una muestra es representativa de una población, es posible inferir importantes conclusiones sobre la población a partir del análisis de la muestra. Ya que dicha inferencia no es del todo exacta, el lenguaje de las probabilidades aparecerá al establecer nuestras conclusiones.

La estadística es más que sólo números: son los datos, lo que se hace con los datos, lo que se aprende de los datos y las conclusiones resultantes. En esta capacitación trabajaremos con la estadística descriptiva.

Glosario:

Población: Es la colección completa de todos los elementos (puntuaciones, personas, mediciones, etcétera) a estudiar. Se dice que la colección es completa, pues incluye a todos los sujetos que se estudiarán.

Muestra: Es un subconjunto de miembros seleccionados de una población.

Muestra representativa (aleatoria): Todos los miembros de la población tienen la misma posibilidad de ser elegido.

Censo: Es la colección de datos de *cada uno* de los miembros de la población.

Datos: Observaciones recolectadas (como mediciones, géneros, respuestas de encuestas).

Parámetro: Medición numérica que describe algunas características de una **población**.

Estadístico: Medición numérica que describe algunas características de una **muestra**.

Datos cuantitativos: Consisten en números que representan conteos o mediciones.

Datos cualitativos (o categóricos o de atributo): Se dividen en diferentes categorías que se distinguen por alguna característica **no numérica**.

Datos discretos: Resultan cuando el número de posibles valores es un número finito, o bien, un número que puede contarse. (Estos es, el número de posibles valores es 0, 1, 2, etcétera).

Datos continuos (numéricos): Resultan de un infinito de posibles valores que pueden asociarse a puntos de alguna escala continua, cubriendo un rango de valores sin huecos ni interrupciones.

Distribución de frecuencias: Lista de valores de datos (ya sea de manera individual o por grupos de intervalos), junto con sus frecuencias (o conteos) correspondientes.

Frecuencia relativa: Se obtiene dividiendo cada frecuencia de clase entre el total de frecuencias.

Distribución de frecuencias relativas: Incluye los mismos límites de clase que una distribución de frecuencias, pero utiliza las frecuencias relativas en lugar de las frecuencias reales.

Distribución de frecuencias acumulativas: La frecuencia acumulativa de una clase es la suma de las frecuencias acumulativas para esa clase y todas las clases previas.

Límites de clase inferiores: Son las cifras más pequeñas que pueden pertenecer a las diferentes clases.

Límites de clase superiores: Son las cifras más grandes que pueden pertenecer a las diferentes clases.

Fronteras de clase: Son las cifras utilizadas para separar las clases, aunque sin los espacios creados por los límites de clase. Se obtienen de la siguiente manera: se determina el tamaño del espacio entre el límite de la clase superior de una clase y el límite de la clase inferior de la siguiente. Se suma la mitad de esa cantidad a cada límite de clase superior, para obtener las fronteras de clase superiores; se resta la mitad de esa cantidad de cada límite de clase inferior, para obtener las fronteras de clase inferiores.



Marca de clase: Son los puntos medios de las clases.

Anchura de clase: Es la diferencia entre dos límites de clase consecutivos o dos fronteras de clase inferiores consecutivas.

Materiales y equipo:

1. Papelotes
2. Marcadores
3. Cinta adhesiva
4. Reglas
5. Calculadoras gráficas *TI-84 Plus*
6. Computadora
7. Proyector digital (*Infocus*)

Proceso Educativo:

- I. Pre y Pos prueba
 1. Se evaluará el conocimiento de los participantes antes de la capacitación con la Preprueba y el conocimiento después con la Posprueba (documentos adjuntos).
- II. Assessment Continuo
 1. Obviamente la preprueba y la posprueba son parte del assessment de la capacitación. Es la primera ayuda al capacitador para tomar decisiones acerca del conocimiento que tiene el participante del tema y de las próximas actividades que llevará a cabo. Mientras la posprueba ayuda al capacitador a tomar decisiones de la necesidad de re enseñanza en próximas capacitaciones.
 2. Las hojas de trabajo, el capacitador las utilizará como assessment. Los participantes estarán cotejando su aprendizaje en la medida que se discutan las mismas en grupo grande. Además, el capacitador las corrige y las utilizarlas para tomar decisiones.
 3. Durante todas las actividades el capacitador estará haciendo observaciones mientras se mueve entre las parejas, cuando los participantes discuten con su pareja y cuando presentan sus respuestas a las preguntas. Esto le permite hacer conclusiones del aprendizaje de éstos y los próximos pasos a seguir.

Primera parte

- I. **Inicio:** *Explorando las concepciones previas*
 1. La actividad está diseñada para indagar el conocimiento que tienen los participantes acerca del concepto encuesta.
 - i. Se dividen los maestros en cuatro grupos y se le pide que en un papelote desarrollen un organizador gráfico alrededor del concepto **distribución de frecuencias**.
 - ii. Cada grupo presenta su organizador gráfico al grupo grande. El capacitador y los participantes no pasarán juicio sobre el organizador gráfico. El capacitador estará observando los organizadores e identificará concepciones erróneas, si las hay, para luego a través de la capacitación hacer énfasis en las mismas y corregirlas. En el cierre de la capacitación los participantes volverán a revisar los organizadores y harán los arreglos pertinentes. De esta forma tendrán la

oportunidad de percatarse de los posibles errores y corregirlos, mientras el capacitador tendrá un assessment final.

II. Desarrollo

Actividad 1: Población vs muestra

1. El capacitador plantea la siguiente situación: **En reportero del canal 2 pretende auscultar, entre las personas mayores, la opinión sobre el trabajo del Gobernador. Se para en una esquina del Viejo San Juan y pregunta a 10 adultos si creen que el Gobernador está haciendo un buen trabajo.** Pregunta, ¿cuál es la población a la que el reportero quiere auscultar? Contestación: todas las personas adultas del país. Pregunta, ¿cuál es la muestra que pudo auscultar? Contestación: las 10 personas encuestadas. Pregunta, ¿la muestra es representativa de la población? Contestación: no.
2. El capacitador define los conceptos población, muestra y muestra representativa (ver Glosario).
3. El capacitador plantea la siguiente situación: **Un sondeo de la firma Gallup preguntó a 1087 adultos estadounidenses: “¿Consume bebidas alcohólicas como licor, vino o cerveza o es abstemio?”.** Pregunta, ¿cuál es la población del sondeo? Contestación: el conjunto de todos los estadounidenses adultos. Pregunta, ¿cuál es la muestra? Contestación: los 1087 sujetos del sondeo. Pregunta, ¿la muestra es representativa? Contestación: no, este tipo de sondeo se hace sin tener en cuenta la aleatoriedad, es muy difícil lograrlo.
4. El capacitador plantea la siguiente situación: **Cada 10 años el gobierno de EU intenta obtener un censo de cada ciudadano.** Pregunta: ¿qué entienden por censo? Permite que los participantes planteen sus respuestas y define el concepto (ver glosario).
5. El capacitador reparte la Hoja de Trabajo 1 (HT1) y discute con los participantes las instrucciones. Mientras los grupos resuelven los problemas el capacitador contesta preguntas y se asegura que los grupos trabajen colaborativamente. (Las contestaciones se encuentran en el documento Hoja de Trabajo 1: Soluciones)
6. El capacitador discute los resultados de la HT1 en grupo grande.

Nota: Medición de la desobediencia

¿De qué manera se recolectan datos que parecen imposibles de medir, como el nivel de desobediencia de las personas? El psicólogo Stanley Milgram ideó el siguiente experimento: un investigador enseñó a un sujeto voluntario a operar un tablero de control que administraba “choques eléctricos” cada vez más dolorosos a una tercera persona. En realidad no se daban tales choques y la tercera persona era un actor. El voluntario iniciaba con 15 voltios y fue instruido para incrementar los choques en aumentos de 15 voltios. El nivel de desobediencia fue el punto donde el sujeto se negaba a incrementar el voltaje. Resultó sorprendente que dos terceras partes de los sujetos obedecieran las órdenes aun cuando el actor gritaba y fingía sufrir un ataque cardiaco.

Obtenido de Triola, M.F. (2004). *Estadística, novena edición*. Pearson, México.

Actividad 2: Tipos de datos

1. El capacitador plantea la siguiente pregunta: ¿Qué entiende cuando hablamos de datos? Permite la discusión de los participantes y guía la discusión para definir el concepto dato (ver el glosario). El capacitador enfatiza que la palabra correcta es **dato** no “data”.
2. El capacitador explica la diferencia entre los conceptos **parámetro** y **estadístico** (ver el glosario). Utiliza los siguientes ejemplos para aclarar los conceptos:
 - a. Cuando Alejandro García Padilla fue elegido gobernador de PR, recibió el 47.73% de 1,877,179 votos (datos de la Comisión Estatal de Elecciones). Si suponemos que el conjunto de todos los votos es la población considerada, entonces el 47.73% es un **parámetro**, no un estadístico.
 - b. Con base en una muestra de 877 ejecutivos encuestados, se encontró que el 45% de ellos no contrataría a alguien con un error ortográfico en su solicitud de empleo. Esta cifra del 45% es un estadístico, ya que está basada en una **muestra**, no en la población completa de todos los ejecutivos.
3. El capacitador explica que los datos se pueden clasificar en datos **cuantitativos** o **cualitativos** (ver el glosario). Utiliza los siguientes ejemplos para aclarar los conceptos y le pide a los participantes que mencionen ejemplos adicionales:
 - a. Los pesos de los participantes de la capacitación.
 - b. El género de los participantes de la capacitación.

El capacitador enfatiza que cuando se trabaja con datos cuantitativos, es importante utilizar las unidades de medida apropiadas, tales como dólares, horas, pies, metros y otras. Debemos ser especialmente cuidadosos para observar aquellas referencias como “todas las cantidades están en *miles de dólares*” o “todos los tiempos están en *centésimas de segundo*” o “las unidades están en *kilogramos*”. Ignorar unidades de medida como éstas podría llevar a conclusiones incorrectas. La NASA perdió su Mars Climate Orbiter de 125 millones de dólares cuando la sonda se estrelló, porque la programación de control tenía los datos de aceleración en unidades *inglesas*, pero ellos incorrectamente consideraron que estaban en unidades *métricas*.

4. El capacitador plantea las siguientes situaciones:
 - a. La cantidad de huevos que pone una gallina.
 - b. La cantidad de leche que las vacas producen.
 - c.

El capacitador pregunta, ¿qué tienen en común estos datos? Contestación, son datos cuantitativos. Pregunta, ¿qué tienen diferente estos datos? Contestación, La cantidad de huevos que ponen las gallinas se pueden contar, son 0, 1, 2, 3, etc. Pero no podemos tener 1.5 huevos, estos datos se dice que son **discretos**. Mientras que la

cantidad de leche que producen las vacas son mediciones que pueden tomar cualquier valor dentro de un intervalo continuo, esto es, en un intervalo de tiempo una vaca producirá una cantidad de leche que puede ser cualquier valor entre 0 y 5 galones. Es posible obtener 2.343115 galones, ya que la vaca no está restringida a producir cantidades discretas de 0, 1, 2, 3, 4, o 5 galones, estos datos son **continuos**. El capacitador define **datos discretos** y **datos continuos** (ver el glosario). El capacitador pide a los participantes que mencionen un ejemplo de datos discretos y un ejemplo de datos continuos.

5. El capacitador reparte la Hoja de Trabajo 2 (HT2) y discute con los participantes las instrucciones. Mientras los grupos resuelven los problemas el capacitador contesta preguntas y se asegura que los grupos trabajen colaborativamente. (Las contestaciones se encuentran en el documento Hoja de Trabajo 2: Soluciones)
6. El capacitador discute los resultados de la HT2 en grupo grande.

Actividad 3: Nivel de medición

Otra forma común de clasificación de los datos es el uso de cuatro niveles de medición: nominal, ordinal, de intervalo y de razón. Cuando la estadística se aplica a problemas reales, el nivel de medición de los datos es un factor importante para determinar el procedimiento a usar. Por ejemplo, no tendría sentido calcular un promedio de números del seguro social, ya que estos números son datos que se usan como identificación, y no representan mediciones ni conteos de algo. Por lo misma razón, no tendría sentido calcular un promedio de los números en las camisetas de los jugadores del equipo nacional de baloncesto.

1. El capacitador pregunta, ¿cuál es el color de su automóvil?, en un papelote hace una lista con las contestaciones. Pregunta, ¿podemos calcular la media de estos datos? El capacitador escucha las contestaciones y explica que estos datos carecen de un orden o de un significado numérico, por tal razón, no pueden utilizarse para realizar cálculos. El capacitador define **nivel de medición nominal** (ver el glosario). El capacitador le pide a los participantes que ofrezcan ejemplos y los discuten.

Nota: A veces se asignan números a las diferentes categorías (en especial cuando los datos se codifican para el uso de sistemas de cómputos), pero tales números no tienen significado computacional y cualquier media aritmética que se calcule carece de sentido. La medida de tendencia central más apropiada para este tipo de dato es la moda.

2. El capacitador pregunta, ¿qué nota obtuviste en el primer curso de inglés que tomaste en la universidad?, en un papelote hace una lista con las contestaciones. Pregunta, ¿podemos ordenar estos datos? El capacitador escucha las respuestas y explica que estos datos pueden ordenarse, sin embargo, no es posible determinar diferencias entre ellas. Esto es, la nota, A, B, C, D, o F, sabemos que A es más alta que B (por lo tanto existe orden), pero no podemos restar B de A (por lo tanto, no se calcula la diferencia). El capacitador define **nivel de medición ordinal** (ver el glosario). El capacitador le pide a los participantes que ofrezcan ejemplos y los discuten.

Nota: Los datos ordinales ofrecen información sobre comparaciones relativas, aunque

no sobre las magnitudes de las diferencias. Por lo general, los datos ordinales no se usan para cálculos como media, pero esta norma se quebranta en ocasiones (como cuando se usan calificaciones con letras para calcular el punto promedio de la calificación).

3. El capacitador pregunta, ¿en qué año naciste?, en un papelote hace una lista con las contestaciones. Pregunta, ¿podemos ordenar estos datos? Contestación, sí. Pregunta, ¿la diferencia entre dos datos tiene significado? Contestación, sí. Pregunta, ¿estos datos son ordinales? Contestación, no. El capacitador explica que estos datos no pueden clasificarse como ordinales pues no cumplen la definición. Pregunta, ¿qué significa el año 0? Contestación, el año 0 es arbitrario en vez de ser un punto de partida de cero natural, que representaría “ausencia de tiempo”. El capacitador define **nivel de medición de intervalo** (ver el glosario).

Nota: En general, los datos en este nivel al no tener un punto de partida natural desde cero (donde nada de la cantidad esté presente) las razones no tienen significado. Por ejemplo, las temperaturas corporales 98.2°F y 98.6°F son ejemplos de datos en este nivel de medición. Tales valores están ordenados, y podemos determinar su diferencia de 0.4°F . Sin embargo, no existe un punto de partida natural. El valor de 0°F quizá parezca un punto de partida, pero es arbitrario y no representa la ausencia total de calor. Como 0°F no es un punto de partida desde cero natural, es erróneo decir que 50°F es dos veces más caliente que 25°F .

4. El capacitador pregunta, ¿cuál es su peso en libras?, en un papelote hace una lista con las contestaciones. Pregunta, ¿podemos ordenar estos datos? Contestación, sí. Pregunta, ¿la diferencia entre dos datos tiene significado? Contestación, sí. Pregunta, ¿qué significa peso 0 libras? Contestación, sí, la ausencia de peso. El capacitador explica que estos datos no pueden clasificarse de intervalo pues no cumplen la definición. El capacitador define **nivel de medición de razón** (ver el glosario).

Nota: Este nivel de medición se denomina de razón porque el punto de partida cero hace que las razones o cocientes tengan significado. Por ejemplo, los precios de los libros de texto universitarios son un ejemplo del nivel de medición de razón. Un costo de \$0 efectivamente representa ningún costo y un libro de \$90 es tres veces más costoso que un libro de \$30.

Entre los cuatro niveles de medición, la mayoría de las dificultades surgen con la distinción entre los niveles de intervalo y de razón. Para hacer menos difícil esta distinción, utilice una sencilla “prueba de razón”: considere dos cantidades en las cuales un número es dos veces el otro y pregúntese si “dos veces” se puede usar para describir correctamente las cantidades. Note que un peso de 200 libras es dos veces más pesado que un peso de 100 libras, pero 50°F no es dos veces más caliente que 25°F , los pesos están en el nivel de razón, mientras que las temperaturas Fahrenheit están en el nivel de intervalo. La siguiente tabla resumen las diferencias entre los cuatro niveles de medición.

Niveles de medición de datos		
Nivel	Resumen	Ejemplo
Nominal	Sólo nombres, categorías o etiquetas. Los datos no pueden acomodarse en un esquema de orden.	Origen de estudiantes universitarios: 500 <i>Cayey</i> } 300 <i>Cidra</i> } <i>Sólo nombres, categorías o etiquetas.</i> 250 <i>Caguas</i> }
Ordinal	Rangos de orden que pueden acomodarse, pero no hay diferencias o carecen de significado.	Automóviles de estudiantes: 750 <i>compacto</i> } 150 <i>mediano</i> } <i>Orden determinado por "compacto, mediano, grande".</i> 50 <i>grandes</i> }
De intervalo	Las diferencias son significativas, pero no hay punto de partida natural y las razones no tienen significado.	Temperatura del Campus: 70°F } 87°F } <i>0°F no es "sin calor". 40°F no es dos veces más caliente que 20°F.</i> 91°F }
De razón	Hay un punto de partida natural y las razones tienen significado.	Distancias de viaje de estudiantes: 5 millas } 10 millas } <i>10 millas es dos veces más lejos que 5 millas.</i> 15 millas }

5. El capacitador reparte la Hoja de Trabajo 3 (HT3) y discute con los participantes las instrucciones. Mientras los grupos resuelven los problemas el capacitador contesta preguntas y se asegura que los grupos trabajen colaborativamente. (Las contestaciones se encuentran en el documento Hoja de Trabajo 3: Soluciones)
6. El capacitador discute los resultados de la HT3 en grupo grande.

Pensamiento Crítico e Interpretación versus Fórmulas

Actualmente los profesores piensan que no es tan importante memorizar fórmulas o realizar cálculos aritméticos complejos a mano. Por el contrario, suelen enfocarse en la obtención de resultados por medio del uso de algún tipo de herramienta tecnológica (calculadoras o programas de cómputo), para después entender, de forma práctica, los resultados a través del pensamiento crítico. Por ejemplo, cuando estudie la *desviación estándar* trate de observar por qué la fórmula clave funciona como una medida de variación, después aprenda a calcular los valores de las desviaciones estándar, pero trabaje realmente en la *comprensión* y la *interpretación* de los valores de la desviación estándar.

Aun cuando en la capacitación se incluye, de forma detallada, los casos para procedimientos importantes, no es necesario conocerlos a la perfección en todas las situaciones. No obstante, recomendamos que, en cada caso, realice cálculos manuales antes de utilizar su calculadora o computadora. Lo anterior hará que su comprensión se incremente y podrá apreciar mejor los resultados obtenidos con las herramientas tecnológicas.

Los métodos que estudiaremos suelen denominarse métodos de **estadística descriptiva**, porque su objetivo es *resumir* o *describir* las características importantes de un conjunto de datos. La materia de estadística tiene dos divisiones generales: *la descriptiva* y *la inferencial*. La **estadística inferencial** trata sobre hacer inferencias (o generalizaciones) acerca de una población a partir de los datos de una muestra. La estadística inferencial está fuera del alcance de esta capacitación.

Actividad 4: Distribución de frecuencias, de frecuencias relativas y frecuencias acumuladas

1. El capacitador explica que cuando se trabaja con conjuntos grandes de datos, con frecuencia es útil organizarlos y resumirlos por medio de la construcción de una tabla que liste los distintos valores posibles de los datos (ya sea de forma individual o por grupos), junto con las frecuencias correspondientes, esto es, el número de veces que ocurren dichos datos. El capacitador aprovecha y define el concepto **distribución de frecuencia** (ver glosario).
2. El capacitador presenta la siguiente tabla:

Distribución de frecuencias de los niveles de cotinina de los fumadores	
Cotinina	Frecuencia
0 – 99	11
100 – 199	12
200 – 299	14
300 – 399	1
400 – 499	2

Explica que la distribución de frecuencias resume los niveles medidos de cotinina de 40 fumadores. La frecuencia de una clase particular es el número de valores originales que caen dentro de esa clase. Por ejemplo, la primera clase de la tabla tiene una frecuencia de 11, lo que significa que 11 de los valores originales de los datos están entre 0 y 99, inclusive. El capacitador aprovecha para definir los siguientes conceptos (ver el glosario):

- i. Límites de clase inferiores (en la tabla anterior son 0, 100, 200, 300 y 400).
- ii. Límites de clase superiores (en la tabla anterior son 99, 199, 299, 399 y 499).
- iii. Fronteras de clase (en la tabla anterior los espacios son exactamente de una unidad, de modo que a los límites de clase superiores se le suma 0.5 y a los límites de clase inferiores se le resta 0.5. Las fronteras de la primera clase son -0.5 y 99.5, en tonto que de la segunda son 99.5 y 199.5, y así sucesivamente).
- iv. Marca de clase (en la tabla anterior son 49.5, 149.5, 249.5, 349.5 y 449.5).
- v. Anchura de clase (en la tabla anterior es 100).

El capacitador enfatiza en la diferencia entre la anchura de la clase y la amplitud de clase (la diferencia en el límite superior e inferior de una clase). En el caso anterior la anchura de clase es 100, mientras la amplitud de clase es 99.

Nota: Las distribuciones de frecuencias se construyen por las siguientes razones:

1. Es posible resumir conjuntos grandes de datos.
2. Se logra cierta comprensión acerca de los datos.
3. Se llega a tener un avance para construir gráficas importantes (tales como histogramas).

3. El capacitador describe el procedimiento básico para construir una distribución de frecuencias a la vez que utiliza los datos de la siguiente tabla para construirla:

Niveles medidos de cotinina* en fumadores											
1	0	131	173	265	210	44	277	32	3	35	112
477	289	227	103	222	149	313	491	130	234	164	198
17	253	87	121	266	290	123	167	250	245	48	86
284	1	208	173								
Niveles medidos de cotinina* en fumadores pasivos											
384	0	69	19	1	0	178	2	13	1	4	0
543	17	1	0	51	0	197	3	0	3	1	45
13	3	1	1	1	0	0	551	2	1	1	1
0	74	1	241								
Niveles medidos de cotinina* en no fumadores y no fumadores pasivos											
0	0	0	0	0	0	0	0	0	0	0	9
0	0	0	0	0	0	244	0	1	0	0	0
90	1	0	309	0	0	0	0	0	0	0	0
0	0	0	0								

* Los datos que se reproducen en la tabla se obtuvieron como parte del National Health and Nutrition Examination Survey, Los valores de los datos corresponden a los niveles medidos de cotinina sérica (en ng/ml) en personas seleccionadas como sujetos de estudio (los datos se redondearon hacia el entero más cercano, de tal modo que un valor de cero no necesariamente implica la ausencia total de cotinina. La cotinina es un metabolito de la nicotina, esto es, es una sustancia que se produce cuando el cuerpo absorbe la nicotina. Porque se sabe que la nicotina se absorbe cuando se consumen cigarrillos, hay una forma indirecta de medir la presencia efectiva del humo del tabaco; esto es, por medio de la cotinina.

- i. Paso 1: Decida el número de clases que desear tener. Debe ser entre 5 y 20, y deben utilizarse números enteros o redondeados. En el ejemplo utilizaremos cinco clases.
- ii. Paso 2: Calcule la anchura de la clase utilizando la siguiente fórmula:

$$AC \approx \frac{(\text{Valor más alto}) - (\text{Valor más bajo})}{\text{Número de clases}}$$

Redondee el resultado para obtener un número más adecuado (generalmente se redondea hacia arriba). Es probable que necesite cambiar el número de clases, pero la prioridad debe ser utilizar valores que sean fáciles de comprender. En el ejemplo el cálculo es 98.2 que se redondea a 100, ya que es un número más conveniente: $AC \approx \frac{491-0}{5} = 98.2 \approx 100$
- iii. Paso 3: Punto de partida: comience por elegir un número para el límite inferior de la primera clase. Elija el valor del dato más bajo o un valor conveniente que sea poco más pequeño. En el ejemplo podemos elegir el 0 como punto de partida, es el valor más bajo de la lista y también porque es un número conveniente.

- iv. Paso 4: Con el uso del límite más bajo de la primera clase y la anchura de la clase, proceda a listar los demás límites de clase inferior. Sume la anchura de clase al punto de partida para obtener el segundo límite de clase inferior. Luego, sume la anchura de clase al segundo límite de clase para obtener el tercero y así sucesivamente. En el ejemplo sume la anchura de clase de 100 al punto de partida de 0 para determinar que el segundo límite inferior de clase es igual a 100. Continúe, y sume la anchura de clase de 100 para obtener los límites inferiores de clase restantes de 200, 300 y 400.
- v. Paso 5: Anote los límites inferiores de clase en una columna vertical y luego proceda a anotar los límites superiores de clase, que pueden identificarse con facilidad. En el ejemplo liste los límites de clase inferiores de forma vertical, como se muestra a continuación:

0 –
100 –
200 –
300 –
400 –

Con esta lista se identifican con facilidad los límites de clase superiores correspondientes, tales como 99, 199, 299, 399 y 499.

- vi. Paso 6: Ponga una marca de clase apropiada para cada dato. Utilice las marcas para obtener la frecuencia total de cada clase. En el ejemplo ponga una marca de clase apropiada a cada dato. Ya que completó las marcas, súmelas para obtener las frecuencias que se presentan en la siguiente tabla:

Distribución de frecuencias de los niveles de cotinina en fumadores	
Cotinina	Frecuencia
0 – 99	11
100 – 199	12
200 – 299	14
300 – 399	1
400 – 499	2

4. El capacitador explica que una variante importante de la distribución básica utiliza las **frecuencias relativas**, que se obtienen fácilmente dividiendo cada frecuencia de clase entre el total de frecuencias. Una **distribución de frecuencias relativas** incluye los mismos límites de clase que una distribución de frecuencias, pero utiliza las frecuencias relativas en vez de las frecuencias reales. Las frecuencias relativas, en ocasiones, se expresan como porcentajes: $frecuencia\ relativa = \frac{frecuencia\ de\ clase}{suma\ de\ todas\ las\ frecuencias}$.
5. El capacitador utiliza los datos de la distribución de frecuencia construida en 3 para construir la distribución de frecuencias relativas de niveles de cotinina en fumadores. La primera clase tiene una frecuencia relativa de $11/40 = 0.275$ o de 27.5%, que se

redondea a 28%. La segunda clase tiene una frecuencia relativa de $12/40 = 0.3$ o 30%, y así sucesivamente. Si se construye de manera correcta, la suma de las frecuencias relativas debe totalizar 1 o 100%, con algunas pequeñas discrepancias, que se permiten al redondear los errores. Puesto que 27.5% se redondeó a 28%, y 2.5% se redondeó a 3%, la suma de frecuencias relativas de la siguiente tabla es de 101%, en vez de 100%.

Distribución de frecuencias relativas de los niveles de cotinina en fumadores	
Cotinina	Frecuencia
0 – 99	28%
100 – 199	30%
200 – 299	35%
300 – 399	3%
400 – 499	5%

6. El capacitador explica que otra variante de la distribución de frecuencias estándar se utiliza cuando se buscan totales acumulativos. La frecuencia acumulativa de una clase es la suma de las frecuencias para esa clase y todas las clases previas. La tabla a continuación muestra la distribución de frecuencias acumulativas del ejemplo anterior.

Distribución de frecuencias acumulativas de los niveles de cotinina en fumadores	
Cotinina	Frecuencia acumulativa
Menos de 100	11
Menos de 200	23
Menos de 300	37
Menos de 400	38
Menos de 500	40

Con el uso de las frecuencias originales de 11, 12, 14, 1 y 2, sumamos $11 + 12$ para obtener la segunda frecuencia acumulativa de 23; después, sumamos $11 + 12 + 14 = 37$, para obtener la tercera, y así sucesivamente. Los límites de las clases fueron reemplazados por expresiones como “menor que”, las cuales describen el nuevo rango de valores.

7. Al comienzo de esta parte presentamos una tabla con los datos de los niveles de cotinina en tres grupos, fumadores, fumadores pasivos y no fumadores ni pasivos. Hay varias preguntas importantes que necesitan contestación: ¿deben preocuparse por su salud las personas que no fuman ante la presencia de fumadores? Para preverlo, en los últimos años las autoridades sanitarias han elaborado muchos reglamentos para restringir el tabaquismo en lugares públicos. ¿Son justificadas dichas regulaciones por razones de salud o sólo provocan dificultades innecesarias a los fumadores. El

capacitador presenta la siguiente tabla de la frecuencias relativas de los tres grupos y le pide a los participantes que la analicen y contesten las preguntas anteriores:

Niveles de cotinina de los tres grupos			
Cotinina	Fumadores	Fumadores pasivos	No fumadores ni pasivos
0 – 99	28%	85%	95%
100 – 199	30%	5%	0%
200 – 299	35%	3%	3%
300 – 399	3%	3%	3%
400 – 499	5%	0%	0%
500 – 599	0%	5%	0%

Es claro que la distribución de frecuencias de los fumadores es muy diferente de las de los otros dos grupos. Debido a que los dos grupos de no fumadores (pasivos y no expuestos) tienen una frecuencia relativa tan alta de cantidades de la primera clase, sería útil comparar más esos conjuntos de datos examinando los valores con mayor detalle.

8. El capacitador reparte la Hoja de Trabajo 4 (HT4) y discute con los participantes las instrucciones. Mientras los grupos resuelven los problemas el capacitador contesta preguntas y se asegura que los grupos trabajen colaborativamente. (Las contestaciones se encuentran en el documento Hoja de Trabajo 4: Soluciones)
9. El capacitador discute los resultados de la HT4 en grupo grande.

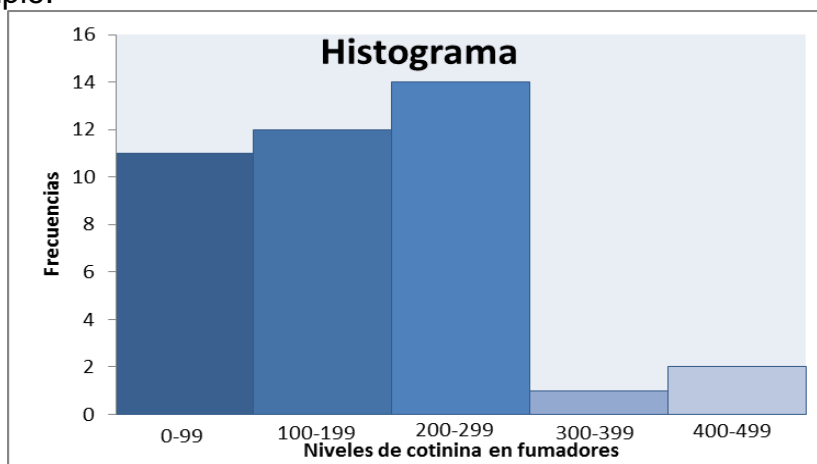
El crecimiento de la estadística

El reportero Richard Rothstein escribió en el *New York Times* que el estudio del álgebra, la trigonometría y la geometría en la escuela preparatoria “deja muy poco espacio para el estudio de la estadística y la probabilidad. Sin embargo, los estudiantes necesitan fundamentos sobre el análisis de datos”. El reportero observó que el cálculo tiene un papel prominente en los estudios universitarios, aun cuando “sólo algunos trabajos, principalmente en áreas técnicas, realmente lo utilizan”. Rothstein citó un estudio realizado por el profesor Clifford Konold, de la Universidad de Massachusetts, quien contó el número de desplegados de datos que aparecen en el *New York Times*. En los ejemplares de 1972, el doctor Konold encontró cuatro gráficas o tablas en cada 10 ediciones semanales (sin incluir las secciones de deportes y negocios), pero en 1982 había ocho, en 1992 fueron 44 y “el próximo año, él podría encontrar más de 100”. El crecimiento de la estadística como disciplina se fomenta, en parte, por el uso creciente de dichos desplegados de datos en los medios de comunicación.

Actividad 5: Visualización de los datos

En la actividad anterior se introdujo la distribución de frecuencias como una herramienta para describir, explorar o comparar distribuciones de conjunto de datos. A continuación estudiaremos el estudio de las distribuciones por medio de la introducción de gráficas, que son dibujos de distribuciones. Las participantes deben considerar que el objetivo no es simplemente la construcción de gráficas, sino más bien aprender la naturaleza de sus distribuciones.

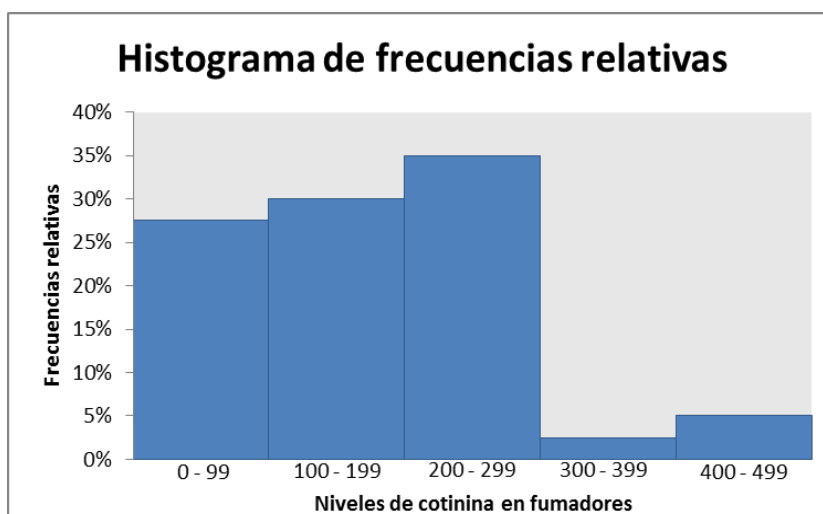
1. El capacitador explica que tras completar una tabla de distribución de frecuencias para un conjunto de datos es posible construir un **histograma** (ver el glosario) y presenta el siguiente ejemplo:



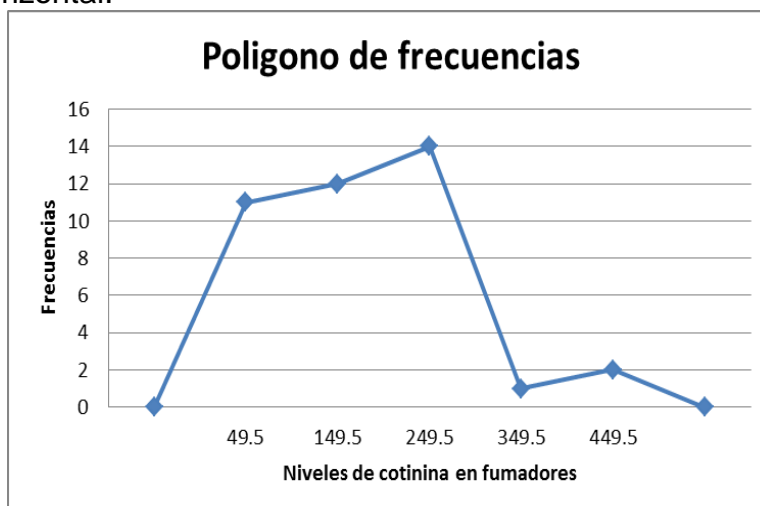
En este se presentan los niveles de cotinina en fumadores, los cuales corresponden a la distribución de frecuencias discutido anteriormente. Cada barra del histograma está marcada con su frontera de clase inferior a la izquierda y su frontera de clase superior a la derecha. En lugar de utilizar fronteras de clase a lo largo de la escala horizontal, suele ser más práctico utilizar los valores de las marcas de clase centradas por debajo de sus barras correspondientes (es lo que utilizaremos en este escrito).

Antes de construir un histograma, a partir de una distribución de frecuencias completa, debemos mencionar algo acerca de las escalas que se utilizan en los ejes vertical y horizontal. La frecuencia máxima (o el siguiente número conveniente más alto) tiene que sugerir un valor para la parte superior de la escala vertical; el cero habrá de colocarse al inicio. La escala horizontal debe subdividirse de modo tal que permita que se ajusten bien todas las clase. De manera ideal, hay que tratar de seguir la regla práctica del intervalo, la cual establece que la altura vertical del histograma debe medir aproximadamente tres cuartas partes de la anchura total. Ambos ejes deben etiquetarse de forma clara.

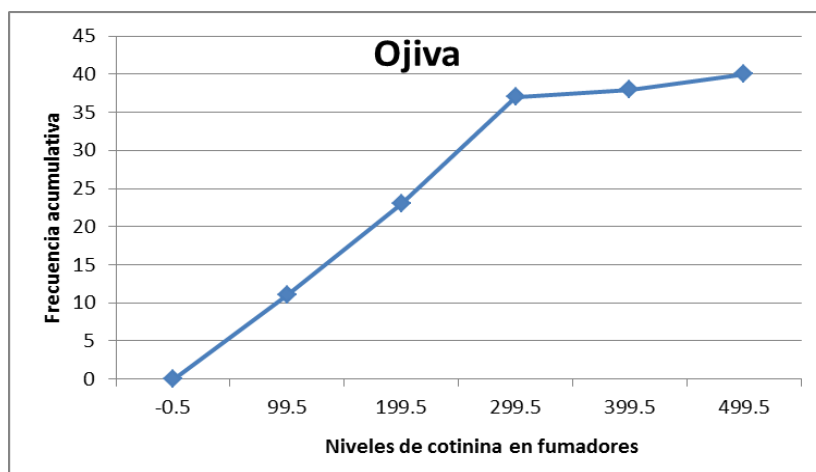
2. El capacitador aclara que el objetivo no es la simple construcción de un histograma, sino aprender algo acerca de los datos. Al analizar el histograma podemos verificar el centro de los datos, la variación, la forma de la distribución y la existencia o ausencia de datos distantes (valores que se encuentran lejos de los demás). Al examinar el histograma anterior notamos que se centra alrededor de 175, que los valores varían aproximadamente desde 0 hasta 500 y que la distribución está cargada hacia la izquierda.
3. El capacitador explica que un **histograma de frecuencias relativas** tiene la misma forma y escala horizontal que un histograma, pero la escala vertical está marcada con las frecuencias relativas en lugar de las frecuencias reales, tal como sucede en la siguiente figura:



4. El capacitador explica que el **polígono de frecuencias** utiliza segmentos lineales conectados a puntos que se localizan directamente por encima de los valores de las **marcas de clase**, como muestra la siguiente figura, los segmentos lineales se extienden hacia la derecha y la izquierda, de manera que la gráfica inicia y termina sobre el eje horizontal:



5. El capacitador explica que una **ojiva** es una gráfica lineal que representa frecuencias **acumulativas**, de la misma forma que la distribución de frecuencias acumulativas es una lista de éstas (remítase a la actividad anterior). La siguiente ojiva corresponde a los frecuencias acumulativas de los niveles de cotinina en fumadores:



El capacitador explica que la ojiva utiliza fronteras de clase, a lo largo de la escala horizontal, y que la gráfica empieza con la frontera inferior de la primera clase, en tanto que finaliza con la frontera superior de la última clase. Las ojivas son útiles para determinar el número de valores que se encuentran por debajo de un valor particular. Por ejemplo, la gráfica muestra que 37 de los valores del nivel de cotinina son menores que 299.5.

6. El capacitador reparte la Hoja de Trabajo 5 (HT5) y discute con los participantes las instrucciones. Mientras los grupos resuelven los problemas el capacitador contesta preguntas y se asegura que los grupos trabajen colaborativamente. (Las contestaciones se encuentran en el documento Hoja de Trabajo 5: Soluciones)
7. El capacitador discute los resultados de la HT5 en grupo grande.

III. Cierre

1. Retomar los organizadores gráficos desarrollados en el inicio para reflexionar acerca de los aprendizajes en la capacitación (*assessment* final). Se le permite a los participantes que vuelvan a los organizadores gráficos y hagan los cambios que sean pertinentes. Se discute con los participantes los cambios propuestos y la razón por qué los hicieron.

Bibliografía:

Departamento de Educación (2011). *Estándares de Contenido y Expectativas de Grado: Programa de Matemáticas*.

Triola, Mario F. (2002). *Estadística*, novena edición. Pearson, México.