

PROMEDIOS, DESVIACIÓN ESTÁNDAR Y SESGOS GUÍA DEL MAESTRO

Autor: Dr. Edwin Morera González

Materia: Matemáticas **Nivel:** 7-9

Conceptos Principales: Medidas de tendencia central y medidas de dispersión.

Conceptos Secundarios: Moda, mediana, media, simetría en conjunto de datos, rango, desviación estándar e interpretación de las medidas de dispersión.

Conocimiento previo: Población, muestra, datos y distribución de frecuencias.

Objetivos específicos:

Durante la actividad el participante:

1. Define los siguientes conceptos:
 - a. Medida de tendencia central
 - b. Moda
 - c. Media
 - d. Mediana
 - e. Media ponderada
 - f. Rango
 - g. Desviación estándar
 - h. Desviación estándar de la muestra
 - i. Distribución simétrica
 - j. Distribución asimétrica
2. Dado un conjunto de datos determina la moda.
3. Dado un conjunto de datos determina la media.
4. Dado un conjunto de datos determina la media ponderada.
5. Dado un conjunto de datos determina la mediana.
6. Dado un conjunto de datos determina la desviación estándar.
7. Determina la mediana en una distribución de frecuencia.
8. Determina la moda en una distribución de frecuencia.
9. Determina la media en una distribución de frecuencia.
10. Dado un conjunto de datos determina el rango.
11. Dado un conjunto de datos determina la desviación estándar.
12. Dado un conjunto de datos determina la desviación estándar de una muestra.
13. Determina la desviación estándar en una distribución de frecuencia.
14. Compara poblaciones a partir de muestras.
15. Resuelve problemas parecidos a los de las PPAA.

**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.

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

Medidas de tendencia central: Valor que se encuentra en el centro o a la mitad de un conjunto de datos. Las más comunes son la media, la moda y la mediana.

Media: La media de un conjunto de datos se obtiene sumando todos los datos y dividiendo luego la suma entre el número de datos. La media de n datos x_1, x_2, \dots, x_n , se calcula como sigue:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Media ponderada: La media ponderada de n números x_1, x_2, \dots, x_n , que se ponderan por sus respectivos factores f_1, f_2, \dots, f_n , se calcula como sigue:

$$\bar{w} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_n \cdot f_n}{f_1 + f_2 + \dots + f_n}$$

Esto es, la media ponderada de un grupo de datos (ponderados) es la suma de todos los productos de los datos por los factores de ponderación, dividida entre la suma de todos los factores de ponderación.

Mediana: La mediana de un conjunto de datos es la medida de tendencia central que implica el valor que está en medio, cuando los valores originales de los datos se presentan en orden de magnitud creciente (o decreciente). La mediana tradicionalmente se denota con \mathcal{M} (se pronuncia “x con tilde”).

Moda: La moda de un conjunto de datos, que tradicionalmente se denota por M es el dato que ocurre con mayor frecuencia.

- Cuando dos valores ocurren con la misma frecuencia y ésta es la más alta, ambos valores son modas, por lo que el conjunto de datos es **bimodal**.
- Cuantos más de dos datos ocurren con la misma frecuencia y ésta es la más alta, todos los valores son modas, por lo que el conjunto de datos es **multimodal**.
- Cuando ningún dato se repite, se dice que no hay moda.

Distribución de datos sesgada: Decimos que una distribución de datos está sesgada si no es simétrica y se extiende más hacia un lado que hacia el otro.

Rango: En cualquier conjunto de datos, el rango del conjunto se define como sigue:

$$\text{Rango} = (\text{valor más grande del conjunto}) - (\text{valor más pequeño del conjunto})$$

Desviación estándar: La desviación estándar de un conjunto de números x_1, x_2, \dots, x_n con media \bar{x} se denota por DE y se define como

$$DE = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Desviación estándar de la población: Dado un conjunto de n números x_1, x_2, \dots, x_n con media \bar{x} , si el conjunto de números representan los datos de una **población** la desviación estándar de la población se denota por la letra griega sigma minúscula (σ), esto es,

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Desviación estándar de la muestra: Dado un conjunto de n números x_1, x_2, \dots, x_n con media \bar{x} , si el conjunto de números representa los datos de una muestra entonces la desviación estándar, s , de la muestra se calcula mediante la siguiente fórmula:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}}$$

Varianza (de un conjunto de valores): Medida de variación igual al cuadrado de la desviación estándar.

Varianza muestral: Cuadrado de la desviación estándar de la muestra s .

Varianza poblacional: Cuadrado de la desviación estándar poblacional σ .

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 los grupos, cuando los participantes discuten en grupo 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 de las medidas de tendencia central y de dispersión.
 - i. Se dividen los maestros en cuatro grupos.
 - ii. El capacitador pregunta el peso de los participantes y los anota en un papelote.
 - iii. Se le pide a cada grupo que, en un papelote, haga una distribución de frecuencia, con los datos anteriores, y que determinen la media, la moda, la mediana, el rango y la desviación estándar.
 - iv. Cada grupo presenta su trabajo al grupo grande. El capacitador y los participantes no pasarán juicio sobre lo presentado. El capacitador estará observando los trabajos 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 resultados de los papelotes 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: Medidas de tendencia central

1. El capacitador define el concepto **medidas de tendencia central** (ver glosario).
2. El capacitador plantea la siguiente situación: **A continuación se incluyen las concentraciones de alcohol en la sangre de conductores que se vieron envueltos en accidentes fatales y que después fueron sentenciados a prisión (con base en datos del Departamento de Justicia de Estados Unidos).**

0.27	0.17	0.17	0.16	0.13	0.24	0.29	0.24
0.14	0.16	0.12	0.16	0.21	0.17	0.18	

Luego pregunta, si las leyes estatales actuales prohíben conducir con niveles por encima de 0.08 o 0.10, ¿están estos niveles significativamente por arriba del máximo permitido?, y permite que los participantes discutan cómo solucionarlo.

3. El capacitador explica que se podrían calcular las medidas de tendencia central de los datos para contestar la pregunta y define la moda, la media y la mediana (ver glosario).
4. El capacitador le pide a los participantes (en grupo) que determinen la media, la moda y la mediana de los datos, y que contesten la pregunta.
5. El capacitador discute en grupo grande la contestación. ($\bar{x} = 0.187$; mediana = 0.170; moda = 0.16. Notemos que las tres medidas de tendencia central están significativamente por arriba del máximo permitido, por lo tanto, la contestación a la pregunta es sí).

6. 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)
7. El capacitador discute los resultados de la HT1 en grupo grande.

El Hombre Promedio

La revista *Men's Health* publicó estadísticas que describen al “hombre promedio”, que tiene 34.4 años de edad, pesa 175 libras, mide cerca de 5 pies 10 pulgadas y se llama Mike Smith. La edad, el peso y la estatura son valores medios, pero el nombre de Mike Smith es la moda que corresponde al nombre y apellidos más comunes. Otra estadística notables es la siguiente: el hombre promedio duerme aproximadamente 6.9 horas por noche, bebe cerca de 3.3 tazas de café al día y consume 1.2 bebidas alcohólicas diariamente; además, gana alrededor de \$36,100 anuales, debe \$2,563 en las tarjetas de crédito y tiene \$3,100 ahorrados en el banco.

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

Actividad 2: La media ponderada

1. El capacitador plantea la siguiente pregunta: Luis estudio un semestre en una universidad de España y obtuvo las siguientes calificaciones:

Curso	Calificación	Unidades
Matemática	A	3
Historia	C	3
Química	C	5
Arte	B	2
Didáctica de la Matemática	A	1

2. El capacitador explica que el concepto *unidades* en las universidades españolas es equivalente al concepto *crédito* en las universidades puertorriqueña. Pregunta: ¿cómo se puede obtener el ***promedio de calificación***? Y permite que los participantes discutan posibles formas de hacerlo.

3. El capacitador explica que un método común es asignarle un valor de 4 puntos a la A, 3 puntos a la B, 2 puntos a la C y 1 punto a la D. Luego se calcula el promedio de calificación de la siguiente forma:

- a. Paso 1: Multiplique el número de unidades de un curso por el número asignado a cada calificación.
- b. Paso 2: Sume estos productos.
- c. Paso 3: Divida entre el número total de unidades.

Curso	Calificación	Pts. De Calificación (P)	Unidades (U)	P · U
Matemática	A	4	3	12
Historia	C	2	3	6
Química	C	2	5	10
Arte	B	3	2	6
Didáctica de la Matemática	A	4	1	4

$$\text{Promedio de puntos de calificación} = \frac{38}{14} \approx 2.71.$$

4. El capacitador explica que el promedio de calificación es ejemplo de una **media ponderada**, porque los puntos de calificación para cada calificación del curso se deben ponderar de acuerdo con el número de unidades del curso (Por ejemplo, cinco unidades de A son mejores que 2 unidades). El número de unidades se conoce como factor de ponderación.
5. El capacitador define media ponderada (ver glosario) y pide a los participantes que resuelvan el siguiente problema: **Luis obtuvo las siguientes puntuaciones en los exámenes y trabajos del curso de Álgebra:**

Examen o trabajo	Puntuación	Por ciento de la nota final
Examen 1	85	20%
Examen 2	92	20%
Trabajo escrito	98	5%
Pruebas cortas	78	10%
Asistencia	90	5%
Examen final	88	40%

Determine el promedio final de Luis en el curso de Álgebra.

6. El capacitador discute la solución del problema anterior. Contestación:

Examen o trabajo	Puntuación (P)	Factor de ponderación (F)	P · F
Examen 1	85	1/5 (20%)	17
Examen 2	92	1/5 (20%)	18.4
Trabajo escrito	98	1/20 (5%)	4.9
Pruebas cortas	78	1/10 (10%)	7.8
Asistencia	90	1/20 (5%)	4.5
Examen final	88	2/5 (40%)	35.2
Total		1 (100%)	87.8

$$\text{Promedio final} = \frac{87.2}{1} = 87.2$$

Aunque usted no lo crea

La revista *Health* comparó las medidas de los maniqués con las medidas de las mujeres. Los siguientes resultados se reportaron como “promedios”, que tal vez representan medias. Estatura de los maniqués: 6 pies; estatura de las mujeres: 5 pies 4 pulgadas. Cintura de los maniqués: 23 pulgadas; cintura de las mujeres: 29 pulgadas. Tamaño de la cadera de los maniqués: 34 pulgadas; tamaño de la cadera de las mujeres: 40 pulgadas. Talla de vestido de los maniqués: 6; talla de vestido de las mujeres: 11. Cuando se camparan las medidas es evidente que los maniqués y las mujeres reales son muy diferentes.

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

Actividad 3: Medidas de tendencia central a partir de distribuciones de frecuencia

1. El capacitador plantea la siguiente situación: **En la compañía de EM Teaching**

Services se pagan los siguiente salarios anuales a sus empleados:

Salario	Números de empleados
\$12,000	8
\$16,000	11
\$18,500	14
\$21,000	9
\$34,000	2
\$50,000	1

2. El capacitador pregunta: ¿Cómo podemos calcular las medidas de tendencia central (media, moda y mediana) de los salarios? Fomenta la discusión entre los participantes asegurándose que concluyan lo siguiente:

- a. Determinar la moda es muy sencillo pues es el dato con frecuencia mayor, \$18,500.
- b. Para determinar la media se utiliza la fórmula de la media ponderada de la siguiente forma:

Salario (S)	Números de empleados (N)	S · N
\$12,000	8	\$96,000
\$16,000	11	\$176,000
\$18,500	14	\$259,000
\$21,000	9	\$189,000
\$34,000	2	\$68,000
\$50,000	1	\$50,000
Total	45	\$838,000

$$\text{Media de salarios} = \frac{\$838,000}{45} = \$18,622 \quad (\text{redondeada})$$

- c. Para determinar la mediana ordenamos los datos, en este caso los salarios ya están ordenados, determinamos el total de las frecuencias y buscamos la posición de la mediana utilizando la siguiente fórmula:

$$\text{Posición de la frecuencia} = \frac{\text{total de las frecuencias} + 1}{2}$$

Si el *total de las frecuencias* es impar, entonces la posición de la frecuencia es la posición de la mediana. Si el *total de las frecuencias* es par, la *posición de la frecuencia* es decimal, a la parte entera de este número se le suma 1 para determinar la posición deseada. En el ejemplo,

$$\text{Posición de la frecuencia} = \frac{45 + 1}{2} = 23, \text{ por lo tanto, el dato en la posición 23 es la mediana, \$18,500.}$$

3. El capacitador explica que tanto la media como la mediana se puede determinar utilizando el menú LIST de la calculadora gráfica cuando tenemos una distribución de frecuencia (Anejo 1).
4. 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)
5. El capacitador discute los resultados de la HT2 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.

Antes de comenzar la **Actividad 4** el capacitador resume lo discutido utilizando la siguiente tabla:

Comparación de la media, mediana y moda						
Medida de tendencia central	Definición	¿Qué tan común es?	Existencia	¿Toma en cuenta cada valor?	¿Se ve afectada por valores extremos?	Ventajas y desventajas
Media	$\bar{x} = \frac{\sum x}{n}$	"Promedio" más conocido	Siempre existe	Sí	Sí	Funciona bien con muchos métodos estadísticos.
Mediana	Valor medio	De uso común	Siempre existe	No	No	Suele ser una buena opción si hay algunos valores extremos
Moda	Valor más frecuente	Se usa en ocasiones	Podría no existir; podría haber más de una	No	No	Apropiada para datos en el nivel nominal.

Actividad 4: Simetría en conjuntos de datos

1. El capacitador explica que la manera más adecuada de analizar un conjunto de datos a menudo depende de si la distribución es simétrica o asimétrica. En una distribución "simétrica", conforme nos desplazamos fuera del punto central, el patrón de frecuencias del lado izquierdo es el mismo (o aproximadamente el mismo) que el del lado derecho. En una distribución "asimétrica", los patrones de la izquierda y de la derecha son diferentes.

2. El capacitador muestra varios tipos de distribuciones simétricas (ver Figura 1) y asimétricas (ver Figura 2). Explica que una distribución de datos está sesgada si no es simétrica y se extiende más hacia un lado que hacia el otro (ver glosario). Pregunta: ¿Cuál de las distribuciones de la Figura 2 son sesgadas?

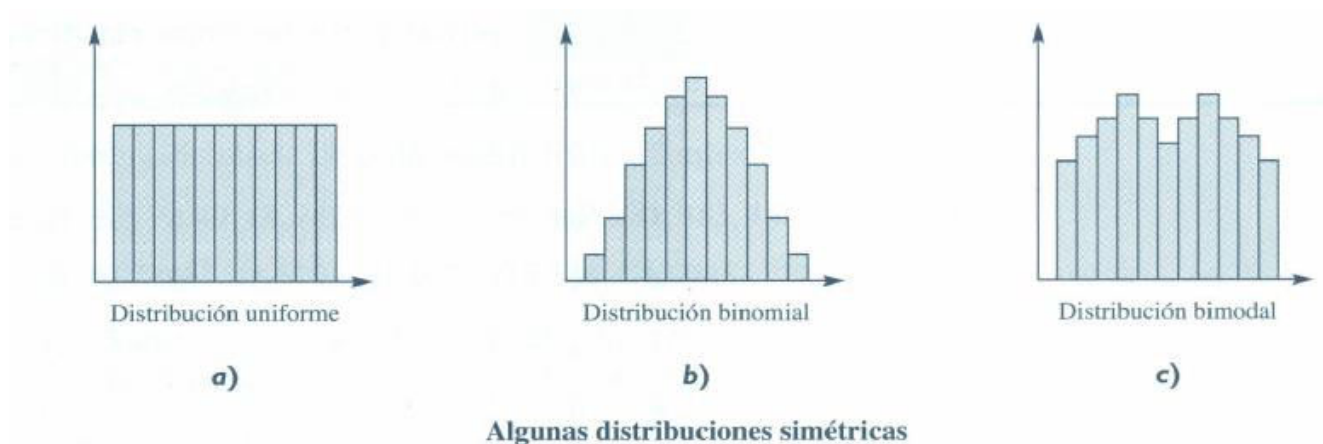


Figura 1

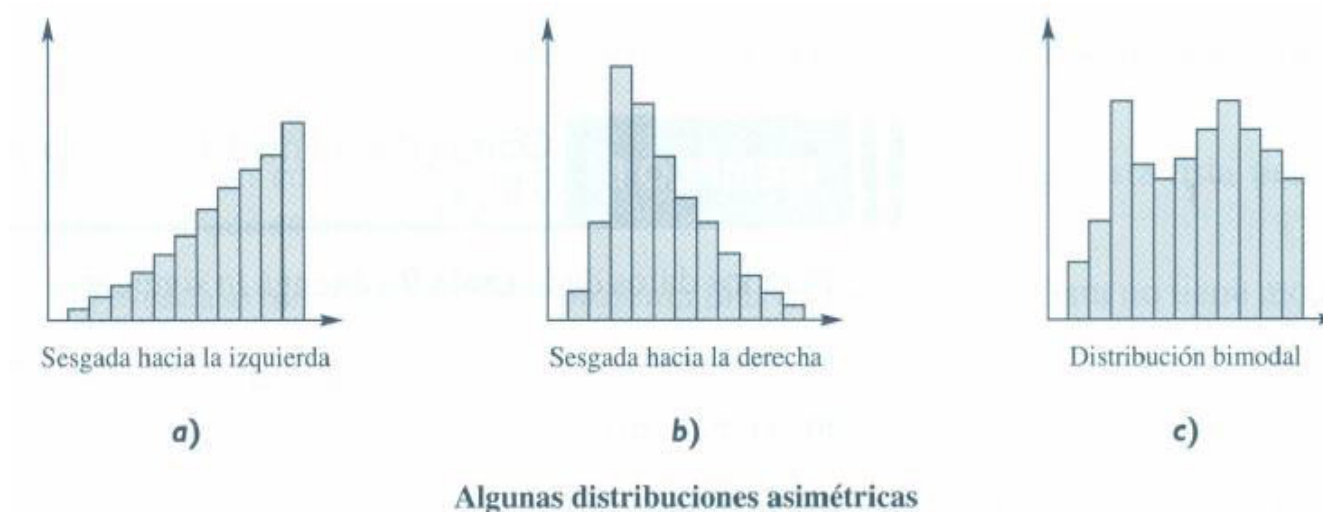


Figura 2

3. Luego de la discusión de los participantes, explica que la gráfica *a)* de la Figura 2 se dice que está sesgada hacia la izquierda y la gráfica *b)* está sesgada hacia la derecha. Mientras la gráfica *c)* no tiene sesgo.
4. El capacitador explica que los datos **sesgados a la izquierda** (que también se denomina como sesgo *negativo*) poseen una cola izquierda más larga, en tanto que la media y la mediana se encuentran a la izquierda de la moda. Aunque no siempre es posible predecirlo, los datos sesgados a la izquierda suelen tener una media menor a la mediana. Mientras los datos **sesgados a la derecha** (que también se denomina como sesgo *positivo*) poseen una cola derecha más larga, en tanto que la media y la mediana se encuentran a la derecha de la moda. Nuevamente, aunque no siempre es posible predecirlo, los datos sesgados a la derecha, suelen tener la media a la derecha de la mediana (ver Figura 3).

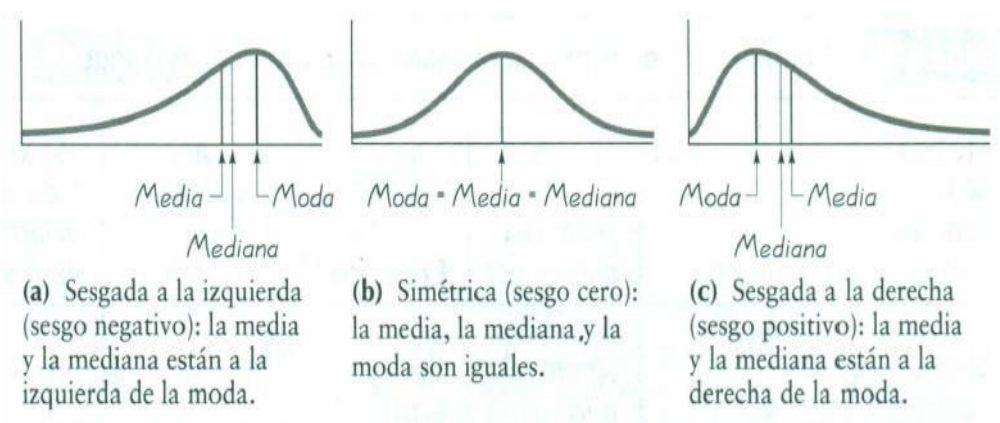


Figura 3

5. El capacitador discute el ejercicio de la PPAA que aparece en el Anejo 2. Hasta el momento los participantes deben saber que una distribución normal es simétrica.

Actividad 5: Medidas de dispersión (rango)

1. El capacitador presenta los siguientes datos y pide a los participantes que comparen la distribución A con la distribución B:

A	B
5	1
6	2
7	7
8	12
9	13

2. El capacitador fomenta la discusión y la dirige para que los participantes observen que aunque ambas distribuciones tienen la misma media y mediana, son bastante diferentes. En la primera, 7 es un valor muy típico (los valores están cerca de él), pero en la segunda, la mayoría de los valores difieren considerablemente de 7. Explica que necesitamos una medida de la dispersión, o *propagación*, de los datos. Pregunta: ¿Conocen alguna medida de dispersión? Esperamos que alguno nos indique el **rango**, pues es muy común, y posiblemente la **desviación estándar**. El capacitador indica que a continuación se discutirán las mismas.
3. El capacitador pregunta: ¿cómo podemos determinar el rango de las distribuciones A y B? Fomenta la discusión para que definan el concepto rango (ver glosario).
4. El capacitador pregunta: ¿cómo podemos describir las distribuciones A y B?, y permite que discutan entre ellos.
5. El capacitador dirige la discusión para que los participantes observen que *aunque las dos distribuciones tienen promedios idénticos, la distribución B presenta tres veces más dispersión, o propagación, que la distribución A.*
6. El rango puede ser engañoso si se interpreta con ligereza. Para que los participantes lo

observen, el capacitador, plantea la siguiente situación: **Revise los puntos**

Proyecto sufragado por el Departamento de Educación mediante el programa: Título I Parte A

alcanzados por Luis y Edwin en cinco exámenes diferentes del curso de matemática, que se muestran a continuación:

Examen	Luis	Edwin
1	28	27
2	22	27
3	21	28
4	26	6
5	18	27
Media	23	23
Mediana	22	27
Rango	10	22

Pregunta: ¿Cuáles calificaciones son más consistentes?

7. El capacitador permite la discusión y la dirige para que los participantes observen que *aunque parece tentador pensar que las calificaciones de Luis son más consistentes que las de Edwin, el rango es menor, las calificaciones de Edwin son más consistentes, con excepción de una calificación mala.* Esa calificación, 6, es un valor atípico que, si se registró correctamente, se debe sin duda a una circunstancia especial.
8. El capacitador dirige la discusión para que los participantes observen que el valor atípico no afecta seriamente la mediana de las calificaciones de Edwin, la cual es más congruente con su rendimiento general que la puntuación de su media.

Actividad 6: Medidas de dispersión (desviación estándar)

Una de las medidas más útiles de dispersión es la ***desviación estándar***, la cual se basa en las desviaciones de los valores de los datos ***respecto de la media***. El capacitador mostrará cómo calcularla de la siguiente manera:

1. Dados los siguientes datos: 32, 41, 47, 53, 57.
 - a. Paso 1: El capacitador calcula las desviaciones respecto de la media, 46, muestra los siguientes datos.

Valor de los datos	32	41	47	53	57
Desviación	-14	-5	1	7	11

Para corroborar se suma las desviaciones, pues esta suma **siempre debe ser igual a 0**.

- b. El capacitador explica que no es posible obtener una medida de la dispersión calculando la media de las desviaciones porque este número siempre es 0, puesto que las desviaciones positivas cancelan exactamente las desviaciones negativas. Para eliminar el problema de números positivos y negativos que se cancelan unos a otros se **eleva al cuadrado** cada desviación, Paso 2.

- c. Paso 2: Elevar al cuadrado las desviaciones.

Valor de los datos	32	41	47	53	57
Desviación	-14	-5	1	7	11
Cuadrado de las desviaciones	196	25	1	49	121

En estos momentos se puede calcular un promedio de los cuadrados de las desviaciones dividiendo su suma entre el número de datos n (5 en este caso), Paso 3.

- d. Paso 3: Promediar los cuadrados de las desviaciones.

$$\bar{x} = \frac{196 + 25 + 1 + 49 + 121}{5} = \frac{392}{5} = 78.4.$$

- e. Paso 4: El promedio que resulta es una medida de dispersión, conocido como varianza: pero una medida común se obtiene calculando la raíz cuadrada de la varianza. Esto compensa, de alguna manera, el cuadrado de la desviación anterior, y da un tipo de promedio de las desviaciones respecto a la media, el

cual se conoce como **desviación estándar de la población**. Esta se denota por la letra griega sigma, σ , minúscula.

$$\sigma = \sqrt{78.4} \approx 8.85$$

2. El capacitador explica la diferencia entre **desviación estándar de la población** y **desviación estándar de la muestra** (ver glosario), y entre **varianza de la población** y **varianza de la muestra** (ver glosario).
3. El capacitador explica cómo se puede utilizar la siguiente tabla para calcular la desviación estándar de la población o de la muestra:

	x	$x - \bar{x}$	$(x - \bar{x})^2$
	1	$1 - 6 = -5$	$(-5)^2 = 25$
	3	$3 - 6 = -3$	$(-3)^2 = 9$
	14	$14 - 6 = 8$	$(8)^2 = 64$
Totales	18		98
$\bar{x} = \frac{18}{3} = 6$			
$\sigma = \sqrt{\frac{98}{3}} = \sqrt{32.67} = 5.72$			
$s = \sqrt{\frac{98}{3-1}} = \sqrt{49} = 7$			

4. El capacitador plantea el siguiente problema y le pide a los participantes que lo resuelvan: **Como parte del National Health Examination se mide el índice de masa corporal en una muestra aleatoria de mujeres. ¿Estará la desviación estándar de la siguiente muestra razonablemente cerca de la desviación estándar de 6.17, que es la desviación de la población?**

19.6, 23.8, 19.6, 29.1, 25.2, 21.4, 22.0, 27.5, 33.5, 20.6, 29.9, 17.7, 24.0, 28.9, 37.7

5. Mientras los participantes discuten en grupos pequeños el capacitador estará haciendo observaciones, mientras se mueve entre los grupos, y contesta preguntas.
6. El capacitador en grupo grande discute la contestación del problema. Contestación:

	x	$x - \bar{x}$	$(x - \bar{x})^2$
	19.6	-5.8	33.64
	23.8	-1.6	2.56
	19.6	-5.8	33.64
	29.1	3.7	13.69
	25.2	-0.2	0.04
	21.4	-4.0	16.00
	22.0	-3.4	11.56
	27.5	2.1	4.41
	33.5	8.1	65.61
	20.6	-4.8	23.04
	29.9	4.5	20.25
	17.7	-7.7	59.29
	24.0	-1.4	1.96
	28.9	3.5	12.25
	37.7	12.3	151.29
Totales	380.5		449.23
$\bar{x} = \frac{380.5}{15} \approx 25.4$		$s = \sqrt{\frac{449.23}{14}} = 5.66$	

La desviación estándar de la muestra no está cerca de la desviación estándar de la población.

7. 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)
8. El capacitador discute los resultados de la HT3 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.

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

Actividad 7: Usando la calculadora gráfica

Como mencionamos anteriormente es posible utilizar la calculadora gráfica para calcular la desviación estándar de la muestra. El capacitador discute el Anejo 3.

III. Cierre

1. Cada grupo retoma los resultados de los papelotes desarrollados en el inicio para reflexionar acerca de los aprendizajes en la capacitación (*assessment* final). Se le permite a los participantes que 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*.

Miller, C.D., Heeren, V.E. y Hornsby, J. (2013). *Matemática: Razonamiento y Aplicaciones*. Pearson, México.

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

