

GUÍA DIDÁCTICA

PROBABILIDAD Y ESTADÍSTICA I



Presentación

El Colegio de Bachilleres de Chiapas a través de la Dirección Académica, la Subdirección de Desarrollo Académico y del Departamento de Formación y Seguimiento a la Academia considera entre sus propósitos, proporcionar a las jóvenes herramientas de apoyo complementario al trabajo desarrollado por el docente para su aprendizaje.

Con la finalidad de contribuir con la formación propedéutica se presenta la guía didáctica de Probabilidad y Estadística I, de manera que el alumno cuente con material didáctico que le permita fortalecer sus conocimientos con base en los contenidos de la asignatura.

La Probabilidad es la parte de las matemáticas encargada del estudio de los fenómenos o experimentos aleatorios; mientras que la Estadística es la recolección de datos con el objetivo de ayudar a rechazar o aceptar afirmaciones, y por lo tanto, permite la toma de decisiones.

La presente guía fue diseñada con el trabajo colegiado de docentes del Colegio de Bachilleres de Chiapas con la misión de ofrecer temas, información y contenido adecuado para su aprendizaje; por lo que el alumno encontrará lecturas, ejemplos y ejercicios para consolidar los conocimientos.

**Colegio de Bachilleres de Chiapas
Dirección Académica
Subdirección de Desarrollo Académico
Depto. de Formación y Seguimiento a la Academia**

Directorio

Dra. Nancy Leticia Hernández Reyes
Directora General

Ing. Luis Alberto Hernández Zambrano
Director Académico

Mtra. María Eunice López Antonio
Subdirectora de Desarrollo Académico

Mtra. Elba D. Casanova Ozuna
Jefa del Depto. de Formación y Seguimiento a la Academia

Guía elaborada por:

Mtro. Raúl Cano Balbuena, docente del plantel 08 Tapachula.
Coordinación de Zona Costa.

Mtro. Ignacio Hidalgo Tirado, docente del CEMSaD 215 Libertad Ventanas.
Coordinación de Zona Costa.

Mtro. Ahmid Rosby Solórzano Osorio, docente del plantel 38 Suclumpá.
Coordinación de Zona Norte

Colaboración especial:

MEH. Sofía Elvira Cordero Santos
Adscrita al Depto. de Formación y Seguimiento a la Academia

Índice	Pág.
Identificación de la guía	5
Bloque 1. Elementos estadísticos	6
Elementos estadísticos	12
Técnicas de recolección de datos	17
Técnicas de muestreo	18
Bloque 2. Descripción gráfica de un conjunto de datos	25
Representación tubular mediante distribución de frecuencias	27
Determinación de clases para datos agrupados	28
Criterios, representación y análisis de gráficas	31
Bloque 3. Medidas estadísticas	37
Medidas de tendencia central para datos simples y agrupados	40
Medidas de dispersión para datos simples y agrupados	52
Límites estadísticos	59
Cuartiles, deciles y percentiles	59
Bloque 4. Comportamiento de dos variables	67
Glosario	91
Fuentes de consulta	96

Identificación de la guía de Probabilidad y Estadística I

Objetivo general

Para Probabilidad y Estadística I se espera que el alumno emplee la estadística como herramienta para organizar, resumir y transmitir información; así como estimar comportamientos de su entorno que le permitan una mejor toma de decisiones, favoreciendo su pensamiento crítico.

Nombre de la asignatura o curso

Probabilidad y Estadística I

Programa al que pertenece

Componente de Formación Propedéutica del Bachillerato General

Semestre o grado al que pertenece

Quinto Semestre

Área a la que pertenece, según el plan de estudios en que se imparte

Matemáticas

Seriación

Probabilidad y estadística II

Valor en créditos: seis

Propedéutica (Optativa u obligatoria)

Obligatoria

Bloque 1. Elementos estadísticos



Introducción al bloque de aprendizaje

En este bloque 1 manejarás las nociones y conceptos básicos de la Estadística, tanto los elementos como las técnicas que permitan el uso de un lenguaje propio de la asignatura en situaciones reales e hipotéticas de su contexto.

Objetivos específicos de aprendizaje del bloque

Aplica crítica y reflexivamente los elementos estadísticos; así como la información recolectada a través de las diferentes técnicas, su tipo de variabilidad y su comportamiento en diferentes contextos.

Competencias genéricas/disciplinares:

Competencias Genéricas:	Competencias Disciplinarias:
<p>3. Elige y practica estilos de vida saludable. CG3.2 Toma decisiones a partir de la valoración de las consecuencias de distintos hábitos de consumo y conductas de riesgo.</p> <p>5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos. CG5.2 Ordena información de acuerdo a categorías, jerarquías y relaciones.</p> <p>8. Participa y colabora de manera efectiva en equipos diversos. CG8.1 Propone maneras de solucionar un problema o desarrollar un proyecto en equipo, definiendo un curso de acción con pasos específicos.</p>	<p>CDEM 5. Analiza las relaciones entre dos o más variables de un proceso social o natural para determinar o estimar su comportamiento.</p> <p>CDEM 7. Elige un enfoque determinista o uno aleatorio para el estudio de un proceso o fenómeno y argumenta su pertinencia.</p>

Interdisciplinariedad/ejes transversales:

Interdisciplinariedad:	Ejes transversales:
<ul style="list-style-type: none">● Geografía● Estructura Socioeconómica de México.● Se retomarán las asignaturas que en cada plantel se impartan en 5o. Semestre, tanto del Componente de Formación Propedéutica como el de Formación para el Trabajo.	<ul style="list-style-type: none">● Eje transversal social● Eje transversal ambiental● Eje transversal de salud.● Eje transversal de habilidades Lectoras.

Conceptos clave:

Población

Muestra

Dato

Variable y sus tipos

Encuesta

Entrevista

Observación

Experimentación

Documental

Muestreo probabilístico

Muestreo no probabilístico.

Bloque I. Elementos estadísticos

Actividad Inicial. Evaluación diagnóstica. Instrucciones: Relaciona ambas columnas colocando en el paréntesis de la derecha la letra que corresponde a la opción correcta.

a) Conjunto de valores que toma la variable que se estudia en cada uno de los individuos de la muestra.	() Población
b) Variables estadísticas que otorgan un resultado representado por un valor numérico.	() Muestra
c) Variable estadística la cual describe cualidades, circunstancias o características de algún	() Estadística descriptiva

tipo de objeto o persona sin hacer uso de números.	
d) Técnica mediante la cual se obtienen generalizaciones, se hacen suposiciones, se infiere y se toman decisiones con base en una información parcial.	() Entrevista
e) Es la disciplina encargada de recopilar, organizar, presentar e interpretar datos convertidos en información.	() Variable
f) Conjunto de todos los elementos que cumplen características que son de interés para efectuar el estudio.	() Estadística inferencial
g) Técnica de recolección de datos que consiste en provocar las situaciones o eventos que darán pie al análisis del comportamiento de los sujetos.	() Variable cuantitativa
h) Técnica de recolección de datos que comprende la interacción verbal directa.	() Datos
i) A cada una de las propiedades o características de los individuos y que no son iguales en cada uno.	() Variable cualitativa
j) Es un subconjunto o porción de una población que cumple con las características de estudio, se compone de objetos, medidas u observaciones elegidas de la población.	() Experimento

Situación Didáctica

Escolaridad en Chiapas

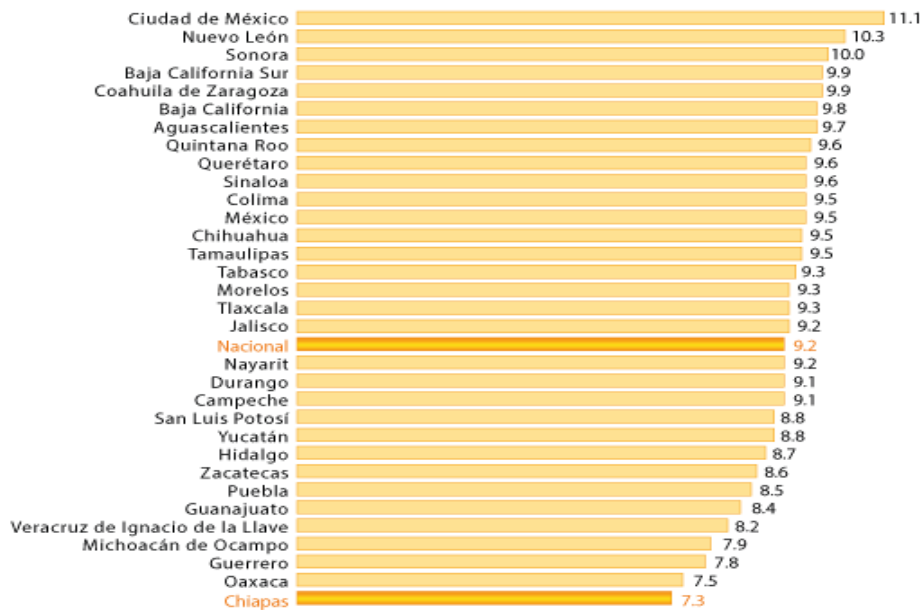
De cada 100 personas de 15 años y más...

14.6	No tienen ningún grado de escolaridad.
57.2	Tienen la educación básica terminada.
16.4	Finalizaron la educación media superior.
11.7	Concluyeron la educación superior.
0.1	No especificado.

En Chiapas, el grado promedio de escolaridad de la población de 15 años y más es de 7.3, lo que equivale a poco más de primer año de secundaria.

A nivel nacional, la población de 15 años y más tiene 9.2 grados de escolaridad en promedio, lo que significa un poco más de la secundaria concluida.

Grado promedio de escolaridad por entidad federativa 2015



Fuente: INEGI. Encuesta Intercensal 2015. Cuéntame, diciembre 15, 2020, sitio web. <http://www.cuentame.inegi.org.mx/monografias/informacion/chis/poblacion/educacion.aspx?tema=me&e=07>

Para comentar en grupo:

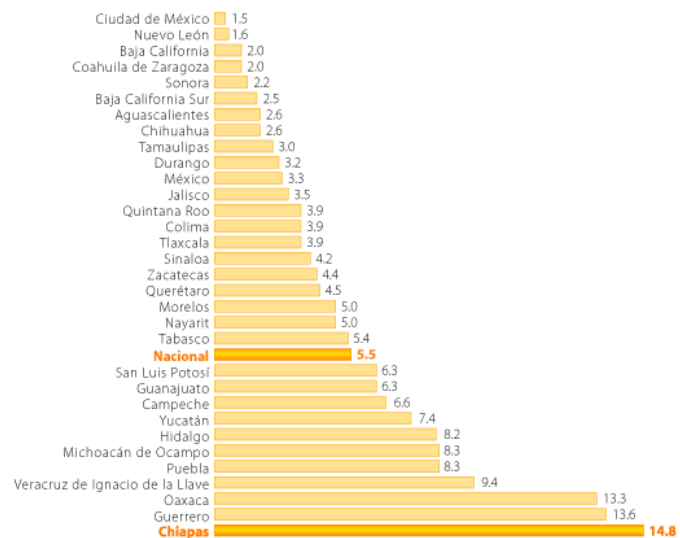
- Según la gráfica, ¿qué entidades tienen la misma escolaridad que Chiapas?
- ¿Qué significa el grado de escolaridad de 7.3 registrado en Chiapas?
- El promedio de escolaridad en Chiapas, ¿es más alto o más bajo que el nacional?
- ¿Qué promedio de escolaridad tiene tu familia?

Analfabetismo

En Chiapas, 15 de cada 100 personas de 15 años y más, no saben leer ni escribir.

A nivel nacional son seis de cada 100 habitantes.

Porcentaje de población analfabeta de 15 años y más por entidad federativa 2015
 Ver Gráfica



Fuente: INEGI. Encuesta Intercensal 2015. Cuéntame, diciembre 15, 2020, sitio web: <http://www.cuentame.inegi.org.mx/monografias/informacion/chis/poblacion/educacion.aspx?tema=me&e=07>

El porcentaje de población analfabeta en Chiapas, ¿es mayor o menor al nacional?

Con base en la información anterior, te darás cuenta que el grado de escolaridad en Chiapas es más bajo que el nacional y que el analfabetismo en Chiapas es mayor que el nacional.

¿Sabes cómo calcular el grado de escolaridad y otras medidas como la media aritmética, la mediana, la moda, desviación estándar y rango aplicado a poblaciones?

Actividad 1

- Ingresa a la página oficial del INEGI www.inegi.org.mx e investiga cuáles son las atribuciones del Instituto y qué fuentes utiliza para generar información estadística básica.
- Investiga el concepto de grado de escolaridad.
- Comenta en el grupo tus conclusiones de la investigación realizada.

Grado de escolaridad: _____

Estadística



Se llama Estadística a la rama de las matemáticas que se sirve de un conjunto de métodos, normas, reglas y principios para la observación, toma, organización, descripción, presentación y análisis del comportamiento de un grupo de datos para la conclusión sobre un experimento o fenómeno.

La estadística tiene básicamente dos divisiones: La estadística descriptiva y la estadística inferencial.

Estadística descriptiva: es la parte de la estadística que estudia las técnicas y métodos que sirven para la observación, tomar organización, descripción, presentación y análisis de datos.

Estadística inferencial: es la parte de la estadística mediante la cual se intenta dar explicación, concluir o inferenciar sobre los experimentos y fenómenos observados mediante la ayuda de la probabilidad, estadística descriptiva y distribución de probabilidad, por lo que resulta una herramienta de suma utilidad para la toma de decisiones.

En este bloque vamos a describir solamente la estadística descriptiva, pero vamos a realizar el siguiente ejercicio para analizar la diferencia entre estos dos tipos de estadística:

A un grupo de 40 alumnos del plantel 13 Tuxtla Oriente del Colegio de Bachilleres de Chiapas le preguntamos “¿cuál es la materia que más les gusta?”. Los datos obtenidos por esta encuesta, en este grupo en particular, es incumbencia de la estadística descriptiva; ya que ordenamos los datos, los analizamos obteniendo sus parámetros como la media, la desviación, los graficamos y hasta lo interpretamos.

Pero si queremos hacer conclusiones a nivel estatal de todos los alumnos del Colegio de Bachilleres de Chiapas, este grupo de 40 alumnos sería tan solo una parte de las diferentes muestras que nos servirían para saber la tendencia de toda la población estudiantil respecto a la materia que más

les gusta, y debemos tomar en cuenta más muestras de estudiantes de otros planteles, por lo cual ya entraríamos en el campo de la estadística inferencial y sus datos deberán analizarse de una manera más profunda, haciendo pruebas de hipótesis para obtener las inferencias o conclusiones a futuro.

Actividad 2. Escribe una conclusión descriptiva y una conclusión inferencial con base en los datos descritos en la siguiente tabla.

Promedio semestral del quinto Semestre			Conclusiones
Grupo	A	B	Descriptiva:
Geografía	7.5	8	
Estructura Socioeconómica de México	8	9	
Cálculo Diferencial	6.5	7	Inferencial:
Temas Selectos de Física I	7.3	7.8	
Probabilidad y Estadística I	8.4	8.8	
Dibujo	8	8.5	

Elementos estadísticos

Estudiar una nueva disciplina científica implica primero aprender su lenguaje, algunos conceptos que debes comprender antes de empezar a estudiar estadística son población, muestra, variables y datos.

Se llama **población** a la cantidad total de cualquier conjunto completo de datos, objetos, individuos o resultados que tengan alguna característica en común que se va a observar o analizar en un problema o experimento.

Se le llama **muestra** a cualquier subconjunto de elementos de la población. El interés de la estadística es proporcionar métodos que permitan elegir una muestra de datos representativos destinados a suministrar información acerca de una población, teniendo como característica fundamental que todos sus elementos deben tener todas las características de la población.

La muestra y sus características dependen del criterio de muestreo empleado para su determinación. Sin embargo, para que una muestra sea representativa de la población, esta deberá contener aproximadamente entre el 5% y el 10% de los datos de la población cuando esta es finita,

además, los elementos de la muestra deben ser escogidos al azar y se deben observar todas las características de la población.

Por **parámetro** se refiere a cualquier valor característico de una población, por ejemplo, el peso promedio, la altura máxima o el estado civil. Este valor es constante.

Por **estadígrafo** o estadístico se refiere a un valor característico obtenido a partir de una muestra. Esta cantidad es variable, puesto que depende de la muestra, ya que de una población se puede elegir un conjunto muy grande de muestras cada una con un valor característico distinto.

Analiza el siguiente ejemplo:

Si se desea conocer qué porcentaje de alumnos y alumnas inscritos en el plantel 08 Tapachula cuentan con teléfono inteligente o tableta, entonces:

Población

Son TODOS los alumnos y alumnas inscritos en el plantel 08 Tapachula.

Muestra

Para este estudio podríamos elegir a los 45 alumnos y alumnas inscritos en el primer Semestre grupo "A".

Actividad 3. Identifica la población y la muestra en cada uno de los casos presentados.

En el plantel 16 Ocosingo se desea saber cuál es el deporte más practicado por los estudiantes. Se realiza una encuesta a 20 alumnos de primer semestre, 15 alumnos de tercer semestre y 15 alumnos de quinto semestre.

Población	Muestra

Se estima que aproximadamente el 80% de los estudiantes del plantel 10 Comitán utilizan el transporte público para trasladarse a sus clases. En una encuesta a 50 alumnos, 40 dijeron utilizar el transporte público.

Población	Muestra

Una compañía de teléfonos desea diseñar nuevos planes de tarifa en el servicio de telefonía móvil en San Cristóbal de Las Casas. Se eligen a 200 usuarios de telefonía móvil y se observan sus gastos en los planes actuales.

Población	Muestra

Datos estadísticos

Los **datos** son las agrupaciones de cualquier número de observaciones relacionadas. Para que se considere un dato estadístico debe tener dos características:

- a) Que sean comparables entre sí,
- b) Que tengan alguna relación.

La recolección de información o recopilación de datos estadísticos se divide en:

Datos internos: Son aquellos datos que no necesitan de observaciones adicionales al experimento, es decir, no es necesario buscar características que proporcionen información adicional acerca del experimento. Ejemplo: las calificaciones de un grupo escolar, un experimento químico, etc.

Datos externos: estos datos pueden ser de dos tipos.

- a) **Datos bibliográficos:** son aquellos ya conocidos y que podemos encontrar fácilmente utilizando bibliografía, registros, actas, etcétera, como los datos históricos, censos y otros.

b) **Datos originales:** son aquellos que podemos obtener mediante métodos de recolección, como las encuestas, plebiscitos, referéndum, y nos proporcionan datos reales y certeros.

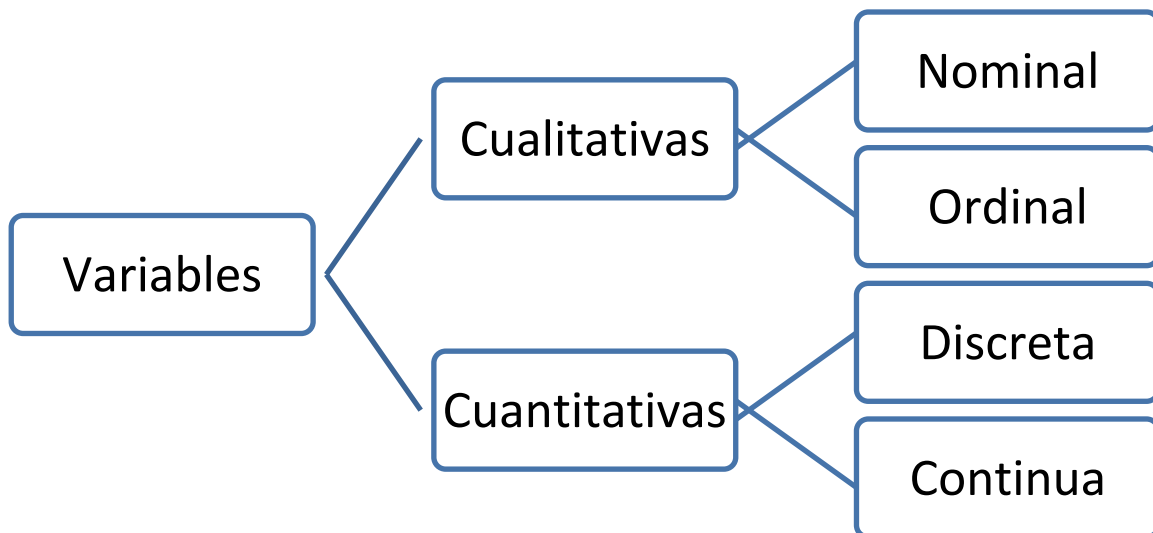
Para **organizar los datos** existen muchas formas de clasificarlos, en general pueden ser determinados de acuerdo a cuatro elementos que son: tiempo, lugar, cantidad y cualidad.

Presentación de datos: después de la organización de los datos, la información se resume en tablas estadísticas con base en arreglos formados de renglones y columnas, adecuados según cronología, geografía, análisis cuantitativo o cualitativo. Los principales elementos de una tabla estadística son: título, unidades, encabezado, cuerpo o contenido, nota de pie y referencias. La información contenida en una tabla estadística también se puede presentar mediante gráficas, siendo las más comunes las de líneas, barras, pictográficas, circulares o de pastel, histograma y polígono de frecuencias.

Variables estadísticas

Una **variable** es una característica de interés sobre cada elemento individual de una población o muestra.

Las variables son las herramientas fundamentales de la estadística y se clasifican de la siguiente manera:



Variables cualitativas: son aquellas que representan atributos de los elementos y no permiten una representación numérica definida. Sin embargo, algunas cualidades pueden ser representadas por códigos numéricos que, en el fondo, generan categorías de orden cualitativo. Ejemplos: estado civil, nacionalidad, lugar de nacimiento, deporte favorito, etc.

Las variables cualitativas a su vez se clasifican en variables nominales y ordinales.

- **Variables nominales:** son variables que no se les puede asignar un orden. Ejemplos: Tipo de sangre, deporte favorito, estado civil, etc.
- **Variables ordinales:** son variables que tienen forma natural de ordenación. Ejemplos: meses del año, estado de salud de una persona según la urgencia del caso, el puesto en una empresa, etc.

Variables cuantitativas: estas variables permiten una escala numérica y las características de los elementos son observados cuantitativamente a través de una medida y escala definida. Ejemplos: edad, longitud, peso, precio, temperatura, etc.

- **Variables discretas:** son aquellas que solo toman valores enteros con rango finito. Ejemplos: cantidad de alumnos en un grupo, la edad de una persona, número de miembros en una familia, etc.
- **Variables continuas:** son aquellas que pueden tomar cualquier valor entre dos valores dados. Es decir, el rango contiene no solo valores enteros sino un intervalo (finito o infinito) de valores reales (esto es, que puede ser fraccionario, decimal o irracional). Ejemplos: altura de una persona, la cantidad de azúcar para endulzar una taza de café, etc.

Actividad 4. Completa la siguiente tabla:

Variable de estudio	Tipo de variable	Posibles valores que puede tomar
Altura de un alumno		
Estado de ánimo de un estudiante		
Temperatura corporal		
Miembros de la familia		
Deporte favorito		

Medios de comunicación		
Meses del año		
Días de la semana		
Promedio semestral		
Materia preferida		
Tipo de sangre		
Nivel de glucosa en la sangre		
Salario mensual		
Número de hermanos		
Consumo de calorías diarias		
Tiempo diario en Internet		
Precio de gasolina		
Escolaridad		
Número de alumnos en el aula		
Género (sexo) de cada estudiante		

Técnicas de recolección de datos

Para obtener información existen varias técnicas, entre éstas: encuesta, entrevista, observación, experimentación y documental.

Encuesta: conjunto de preguntas tipificadas dirigidas a una muestra representativa, para averiguar estados de opinión o diversas cuestiones de hecho. A diferencia de un censo donde todos los miembros de la población son estudiados, las encuestas recogen información de una porción de la población de interés, dependiendo del tamaño de la muestra en el propósito de estudio.

Entrevista: se utiliza para recabar información en forma verbal a través de preguntas que propone el analista. Quienes responden pueden ser gerentes o empleados, los cuales son usuarios actuales del sistema existente, usuarios potenciales del sistema propuesto o aquellos que proporcionarán datos o serán afectados por la aplicación propuesta. El analista puede entrevistar al personal en forma individual o en grupos. Sin embargo, las entrevistas no siempre son la mejor fuente de datos de aplicación. En otras palabras, la entrevista es un intercambio de información que se efectúa cara a cara. Es un canal de comunicación entre el analista y la organización; sirve para obtener información acerca de las necesidades y la manera de satisfacerlas. Por otra parte, la entrevista ofrece al analista una excelente oportunidad para establecer

una corriente de simpatía con el personal usuario, lo cual es fundamental en el transcurso del estudio.

Observación: otra técnica útil para el analista en su progreso de investigación, consiste en observar a las personas cuando efectúan su trabajo. Como técnica de investigación, la observación tiene amplia aceptación científica. Los sociólogos, psicólogos e ingenieros industriales utilizan extensamente esta técnica con el fin de estudiar a las personas en sus actividades de grupo y como miembros de la organización. El propósito de la organización es múltiple: permite al analista determinar que se está haciendo, cómo se está haciendo, quién lo hace, cuándo se lleva a cabo, cuánto tiempo toma, dónde se hace y por qué se hace.

Experimentación: suele mostrarse como una técnica de recolección de datos, sin embargo, la experimentación no es en absoluto una forma de recolección de datos, sino más bien, una estrategia de investigación. Es apropiada cada vez que intentamos determinar el efecto o impacto de una cosa en otra.

Documental: la recolección de datos documental es un instrumento o técnica de investigación general cuya finalidad es obtener datos e información a partir de fuentes documentales con el fin de ser utilizados dentro de los límites de una investigación en concreto. Ninguna guía de recolección de datos puede suministrar una orientación detallada del material a recopilar indicando que documentos son importantes y cuales no lo son, ello depende de las habilidades del investigador, de su experiencia y capacidad para descubrir los indicios que permitan ubicarlos.

Actividad 5

Instrucciones: consulta el link del video de YouTube o bien, recurre a materiales adicionales como libros de texto, videos, páginas Web y realiza un mapa mental sobre las técnicas de recolección de datos.

<https://www.youtube.com/watch?v=1gM3QPRTbFQ>

Técnicas de muestreo

El muestreo es utilizado en la estadística con la finalidad de optimizar recursos (tiempo, mano de obra, materiales e insumos). Para lograrlo, los

métodos estadísticos ofrecen dos tipos de muestreo: probabilístico y no probabilístico.

Técnica de muestreo probabilístico

Los métodos de muestreo probabilístico buscan que todos los elementos que conforman la población tengan igual probabilidad al ser seleccionados en la muestra. Es decir, todas las muestras de un tamaño determinado que se puedan sacar de una población tendrán la misma probabilidad de ser elegidas. Este método de muestreo es el más recomendable, dado que aseguran de alguna manera la representatividad de la muestra que se extrae; se dice de alguna manera porque pueden emplearse técnicas para determinar el tamaño representativo de la muestra para poblaciones finitas e infinitas, a partir del nivel de confianza, error de muestreo y tamaño de la población. A continuación, se describen los principales métodos de muestreo probabilístico.

Muestreo aleatorio simple

Es el más utilizado por su agilidad, sin embargo, es poco útil cuando se tienen poblaciones muy grandes. Los pasos a seguir en este tipo de muestreo es el siguiente: se asigna un número a cada elemento de la población y, por medio de un mecanismo como pelotas dentro de un recipiente o el uso de números aleatorios, se elige el número de elementos requeridos para la muestra.

Muestreo sistemático

Se utiliza cuando el universo es de gran tamaño o ha de extenderse en el tiempo y requiere de una selección aleatoria inicial de observaciones seguida de otra selección de observaciones, obtenida mediante una constante denominada constante de sistematización $C_s = N/n$ donde N es el tamaño de la población y n el tamaño de la muestra. Esta constante nos sirve para determinar cada cuántos elementos o cada cuánto tiempo se debe elegir el siguiente; para ello hay que elegir al azar un número entre 1 y C_s ; de ahí en adelante tomar uno de cada K a intervalos regulares. Ocasionalmente es conveniente tener en cuenta la periodicidad del fenómeno.

Muestreo estratificado

En este muestreo se involucra la división previa de la población en subgrupos, clases o estratos que se suponen más homogéneos, y a los cuales se les asigna una cuota que determina el número de miembros del

estrato que compondrán la muestra, estos son escogidos mediante muestreo aleatorio simple.

Según la cantidad de elementos de la muestra que se han de elegir de cada uno de los estratos, existen dos técnicas de muestreo estratificado:

- a) **Asignación proporcional:** el tamaño de cada estrato en la muestra es proporcional a su tamaño en la población.
- b) **Asignación óptima:** la muestra recogerá más individuos de aquellos estratos que tengan más variabilidad. Para ello, es necesario un conocimiento previo de la población.

Muestreo por conglomerados

Se le llama muestreo por conglomerados al dividir primero la población en grupos o conglomerados convenientes para el muestreo, seleccionando de cada uno de ellos una porción, al azar o por un método sistemático. Bajo este método, aunque no todos los grupos son muestreados, cada grupo tiene una igual probabilidad de ser seleccionado. Por tanto, la muestra es aleatoria. Una muestra por conglomerados, usualmente produce un mayor error muestral que una muestra aleatoria simple del mismo tamaño; sin embargo, puede ser obtenida dentro de un corto periodo de tiempo y a bajo costo. Además, una muestra por conglomerados ofrece la misma precisión en la estimación que una muestra aleatoria simple, si la variación de los elementos individuales dentro de cada conglomerado es proporcionalmente tan grande como la de la población.

Técnica de muestreo no probabilístico

Para algunos tipos de estudios, el muestreo probabilístico puede llegar a generar altos costos y es en estos casos donde se acude a métodos no probabilísticos. Pese a que estos métodos no permiten generalización, porque todos los elementos de la población no poseen la misma probabilidad de ser seleccionados, es posible elegir las unidades muestrales bajo determinados criterios buscando mayor representatividad en la muestra. Los métodos más usados en el muestreo no probabilístico son:

Muestreo por cuotas

También se conoce como muestreo accidental, por ser muy utilizado en las encuestas de opinión. Tiene como premisa el conocimiento amplio de los estratos o grupos relevantes de la población, es por eso que se asemeja al muestreo estratificado, pero difiere al no tener el carácter aleatorio.

Muestreo intencional u opinático

Este tipo de muestreo se utiliza cuando la población es desconocida y se requieren obtener datos representativos. Para ello, se incluyen en la muestra grupos típicos que permitan suministrar la información requerida, por ejemplo, sondeos preelectorales o gustos sobre un evento.

Muestreo bola de nieve

Este tipo de muestreo es utilizado en estudios con poblaciones en las cuales no se conocen los elementos que la integran, pero se sabe que existen. En este proceso se localizan algunos individuos que permitan la vinculación con otros, y así hasta obtener una muestra representativa, por ejemplo: estudios de graduados de una institución, de personas que han cometido un determinado delito, de miembros de una secta religiosa, entre otros.

Actividad 6 En cada uno de los siguientes enunciados determina cual tipo de muestreo es más recomendable aplicar y justifica las razones por las cuales se recomienda:

1. Un docente se encuentra realizando una investigación en una universidad, para constituir la muestra usa estudiantes voluntarios de la Institución.

2. Se desea seleccionar una muestra de 100 alumnos de un total de 1500 estudiantes del Colegio de Bachilleres que se encuentran inscritos en los 13 planteles de la zona Sierra Fronteriza de Chiapas, con el objeto de hacerles una encuesta sobre los servicios educativos que reciben. El número de estudiantes en cada plantel es diferente.

3. Se presentó una lista de 500 tiendas de abarrotes en una ciudad, se elige uno al azar y a partir de esta primera selección, y contando de 20 en 20, se ha ido seleccionando una muestra de 30 tiendas de abarrotes.

Autoevaluación

Lee detenidamente las preguntas y responde colocando una X en el nivel de avance que consideras has logrado a lo largo del bloque 1.

Interpretación del nivel de avance

10% → Excelente, logré el aprendizaje de manera independiente.

8% → Bueno, requerí apoyo para construir mi aprendizaje.

6% → Regular, fue difícil el proceso de aprendizaje y lo logré parcialmente.

5% o menos → Insuficiente, no logré el aprendizaje.

Contenidos	Nivel de avance			
	10%	8%	6%	5% o menos
1. Identificas para qué sirve la estadística en tu vida cotidiana.				
2. Comprendes la diferencia entre estadística descriptiva y estadística inferencial.				
3. Comprendes los conceptos de población y muestra.				
4. Aplicas correctamente los conceptos de población y muestra en problemas cotidianos.				
5. Identificas los diferentes tipos de variables.				
6. Reconoces el concepto				

de dato estadístico.				
7. Sabes diferenciar los tipos de recolección de datos.				
8. Comprendes las técnicas de muestreo probabilístico				
9. Comprendes las técnicas de muestreo no probabilístico.				
10. Aplicas correctamente las técnicas de muestreo en problemas cotidianos.				
Total				

Orientaciones para el estudio

Hoy en día, los jóvenes de Educación Media Superior (EMS) transitan hacia la vida adulta, interactúan en un mundo que evoluciona de la sociedad del conocimiento hacia la sociedad del aprendizaje y la innovación (Joseph Stiglitz, 2014; Ken Robinson, 2015; Richard Gerver, 2013; y Marc Prensky, 2015; entre otros); procesan enormes cantidades de información a gran velocidad y comprenden y utilizan, de manera simultánea, la tecnología que forma parte de su entorno cotidiano y es relevante para sus intereses.

Por lo anterior, en la EMS debe superarse la desconexión existente entre el currículo, la escuela y los alumnos, ya que la misma puede producir la desvinculación educativa de éstos, lo cual, incluso puede derivar en problemas educativos como los bajos resultados, la reprobación y el abandono escolar. Para ello, en primer lugar, hay que entender que los jóvenes poseen distintos perfiles y habilidades (no son un grupo homogéneo) que requieren potenciar para desarrollar el pensamiento analítico, crítico, reflexivo, sintético y creativo, en oposición al esquema que apunte sólo a la memorización; esto implica superar, asimismo, los esquemas de evaluación que dejan rezagados a muchos alumnos y que no miden el desarrollo gradual de los aprendizajes y competencias para responder con éxito al dinamismo actual que los jóvenes requieren enfrentar para superar los retos del presente y del futuro. En segundo lugar, se requiere un currículo pertinente y dinámico, en lugar del vigente que es segmentado y limitado por campo disciplinar, que se centre en la juventud y su aprendizaje, y que ponga énfasis en que ellos son los propios arquitectos de sus aprendizajes.

La escuela, en consecuencia, requiere transformarse de fondo para lograr incorporar en el aula y en la práctica docente las nuevas formas en que los jóvenes aprenden, y lo seguirán haciendo (Gerver, 2013; Prensky, 2013); de no hacerlo, quedará cada día más relegada de la realidad.

Para el estudiante:

Considera invertir varias horas a la semana al estudio de Probabilidad y Estadística I de manera autogestiva.

Recuerda que puedes utilizar otros recursos para tu formación: videos, lecturas, podcasts, o consultar directamente con tu profesor o profesora.

Página sugerida para consulta sobre técnicas de recolección de datos: <http://www.postgradoune.edu.pe/pdf/documentos-academicos/ciencias-de-la-educacion/23.pdf>

Recursos didácticos

Computadora

Pizarrón

Hoja de cálculo.

Evaluación

Tema integrador del Bloque 1:

El docente formará equipos de trabajo de tres personas, el trabajo por cada grupo será encuestar a un número de personas de su comunidad, en el caso de la escuela un grupo, por ejemplo, o hacerlo en la hora de receso entrevistando a personas de manera aleatoria.

El tema a tratar debe ser común como, por ejemplo:

- Hacer una estadística de los equipos favoritos del futbol mexicano y europeo, dando tres lugares más importantes.
- También puede ser del color favorito.
- Así también de la comida favorita de la cocina de la escuela.
- El gasto diario de los alumnos, etc.

El equipo puede hacer una gráfica de barra o pastel donde demuestre sus encuestas y debe presentarlo en una plenaria en el salón.

Se hace una rúbrica donde califique: las encuestas, de temas cualitativos, la presentación con papel bond, proyector, y la forma como explica su encuesta con los resultados.

Bloque 2. Descripción gráfica de un conjunto de datos



Introducción al bloque de aprendizaje

Las descripciones de los conjuntos de datos, nos proporcionan métodos gráficos y tabulares para organizar y resumir datos, para luego usar dicha información y sacar conclusiones, de modo que,

si queremos determinar su significación, rápidamente se pueda tener una idea de las características esenciales de los datos.

Debemos abordar los temas desde el punto de vista de la estadística para que podamos hacer generalizaciones y sacar conclusiones de una población a partir de una muestra representativa de la población.

Objetivos específicos de aprendizaje del bloque

Al término del presente bloque, el alumno tendrá la capacidad de representar de manera gráfica un conjunto de datos, ya sean datos sin agrupar o datos agrupados; para poder hacer inferencias acerca de una población dada.

Competencias Genéricas/Disiplinares

Competencias Genéricas:

- ✓ 4.1 Expresa ideas y conceptos mediante representaciones Lingüísticas, Matemáticas o Gráficas.
- ✓ 5.2 Ordena información de acuerdo a categorías, jerarquías y relaciones.
- ✓ 8.2 Aporta puntos de vista y considera los de otras personas de manera reflexiva.

Competencias disciplinares:

- ✓ CDEM 2 Formula y resuelve problemas matemáticos aplicando diferentes enfoques.
- ✓ CDEM 4 Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos y analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las Tecnologías de Información y la Comunicación.

Interdisciplinariedad/Ejes transversales

Interdisciplinariedad: Geografía, Estructura Socioeconómica de México.

Eje transversal social: se sugiere retomar temas relacionados con la educación financiera, moral y cívica, para la paz (Derecho Humanos), equidad de género, interculturalidad, lenguaje no sexista, vialidad, entre otros.

Eje transversal ambiental: se recomienda abordar temas referentes al respeto a la naturaleza, uso de recursos naturales, desarrollo sustentable, reciclaje, entre otras.

Eje transversal de salud: se sugiere abordar temas relacionados con la educación sexual integral y reproductiva, cuidado de la salud, prevención y consumo de sustancias tóxicas, entre otras.

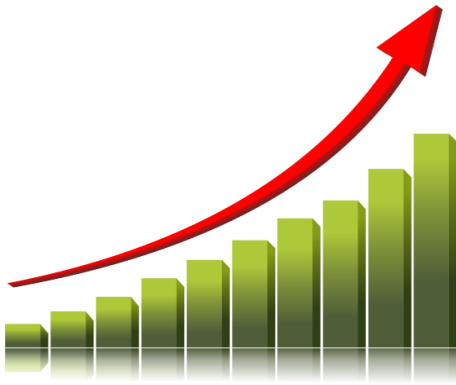
Eje transversal de habilidades lectoras: se recomienda retomar temas relacionados con la lectura, comprensión lectora, lecto-escritura y lectura de textos comunitarios o en lenguas nativas, entre otros.

Conceptos clave:

Frecuencia (absoluta, relativa, acumulada)

Histograma

Polígono de frecuencias



En este bloque necesitas aprender los siguientes términos y su forma de representación ya que debes ser capaz de realizar la descripción gráfica, y ésta es esencialmente la visualización gráfica de la información. Debemos tener presente que la descripción gráfica tiene como objetivo ofrecer una visión de conjunto de datos estadísticos para hacerlo de manera más directa y perceptible.

Frecuencia absoluta

La frecuencia absoluta es el número de veces que aparece un determinado valor en un estudio estadístico; es útil para saber acerca de las características de una población o muestra. Esta medida se puede utilizar con variables cualitativas o cuantitativas siempre que estas se puedan ordenar.

“La frecuencia absoluta se puede usar para variables discretas (las variables se ordenan de menor a mayor) y para variables continuas (las variables se ordenan de menor a mayor agrupadas por intervalos). La frecuencia absoluta se utiliza para calcular la frecuencia relativa. La suma de las frecuencias absolutas es igual al número total de datos de la muestra o población” (Marco San Juan, Francisco. 2021. En línea).

Frecuencia relativa

La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. La frecuencia es igual a la unidad. Es decir, la frecuencia relativa se calcula como el cociente de la frecuencia absoluta de algún valor de la población/muestra (f_i) entre el total de valores que componen la población/muestra (N).

Frecuencia acumulada

La frecuencia acumulada es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.

Francisco Marco San Juan (2021) señala: “para calcular la frecuencia acumulada hay que ordenar los datos de menor a mayor. Para un cálculo más sencillo y una imagen más visual, estos se colocan en una tabla. Tras tener los datos ordenados y tabulados, la frecuencia acumulada se obtiene simplemente de ir sumando una clase o grupo de la muestra con la anterior (primer grupo + segundo grupo, primer grupo + segundo grupo +

tercer grupo y así sucesivamente hasta llegar a acumular del primer grupo al último).”

Regla de Sturges para datos agrupados

“Esta regla es un criterio utilizado para determinar el número de clases o intervalos que son necesarios para representar gráficamente un conjunto de datos estadísticos” (Lifeder.com. 2018). Su utilidad en la Estadística – especialmente, estadística descriptiva- se observa en la construcción de histogramas de frecuencia. La regla de Sturges puede ser representada así:

$$K= 1+ 3,322 * \log_{10} (N)$$

En esta expresión:

- k es el número de clases.
- N es el número total de observaciones de la muestra.
- Log es el logaritmo común de base 10.

Por ejemplo, para elaborar un histograma de frecuencia y conocer el número de intervalos o clases que exprese una muestra aleatoria de la estatura de 142 jóvenes, se realizaría la siguiente operación:

$$k=1 + 3,322 * \log_{10} (N)$$

$$k=1+3,322* \log (142)$$

$$k=1+3,322* 2,1523$$

$$k=8,14 \approx 8$$

Así, la distribución será en 8 intervalos. El número de intervalos siempre debe estar representado por números enteros.

Observa el siguiente ejemplo tomado de matemovil.com (2021).

Una tienda en línea registra el tiempo que tarda la empresa de correos en hacer llegar su mercadería a los clientes. Los tiempos en días registrados son los siguientes:

2	7	10	16	19
22	6	25	5	20

13	32	13	29	18
20	13	6	12	35

Con los datos obtenidos, elaborar una tabla de frecuencias.

Solución:

Calculamos el rango, número de intervalos usando la regla de Sturges, y la amplitud:

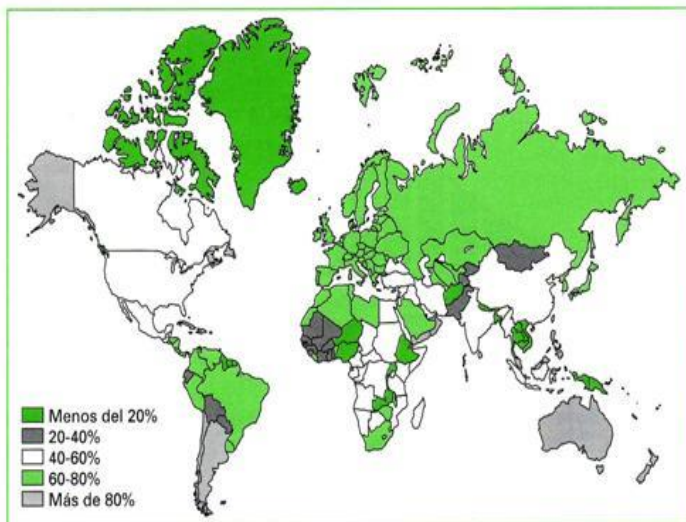
1. Rango: $R = x_{max} - x_{min} = 35 - 2 = 33$

2. Número de intervalos: $k = 1 + 3,322 \log 20 = 5,32 \cong 5$

3. Amplitud: $A = \frac{R}{K} = \frac{33}{5} = 6,6 \cong 7$

La tabla de frecuencias sería la siguiente:

Tiempo	Marca de clase	Frecuencia absoluta	Frecuencia acumulada	Frecuencia relativa	Frecuencia relativa acumulada
[0 - 7)	3.5	4	4	0.2	0.2
[7 - 14)	10.5	6	10	0.3	0.5
[14 - 21)	17.5	5	15	0.25	0.75
[21 - 28)	24.5	2	17	0.1	0.85
[28 - 35]	31.5	3	20	0.15	1
Total		20		1	



Cartograma

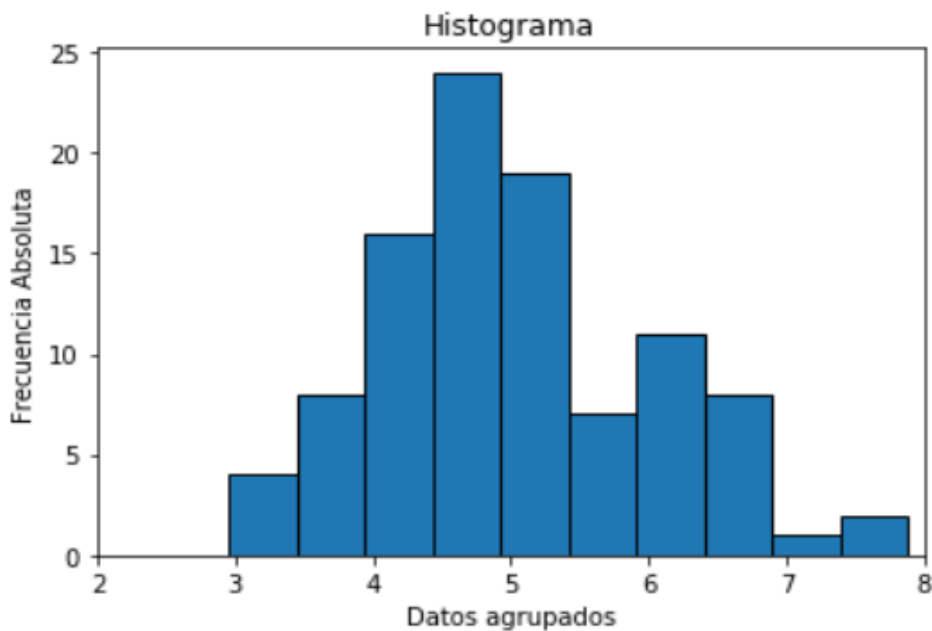
Los cartogramas son gráficos realizados sobre mapas en los que aparecen indicados sobre las distintas zonas, cantidades o colores de acuerdo con el carácter que representan.

En el cartograma de ejemplo observamos la urbanización en el mundo atendiendo a la industrialización.

Histograma

Es la representación gráfica de las frecuencias agrupadas de una variable continua sobre intervalos. A diferencia de los diagramas de barras, los histogramas dibujan rectángulos unidos entre sí, lo que significa que existe continuidad en la variable cuyos valores se representan en el eje horizontal que se halla dividido en intervalos de igual amplitud. Guillermo Westreicher (2021. En línea) señala: "el histograma es entonces un gráfico que permite mostrar cómo se distribuyen los datos de una muestra estadística o de una población. Esto, respecto a alguna variable numérica. En el histograma se suelen usar barras, cuya altura dependerá de la frecuencia de los datos que corresponde al eje Y. En tanto, en el eje X podemos observar la variable de estudio."

Para esto, debemos recordar que en estadística la frecuencia es la cantidad de veces que se repite un suceso.



Fuente: Superprof. Material didáctico (2019)

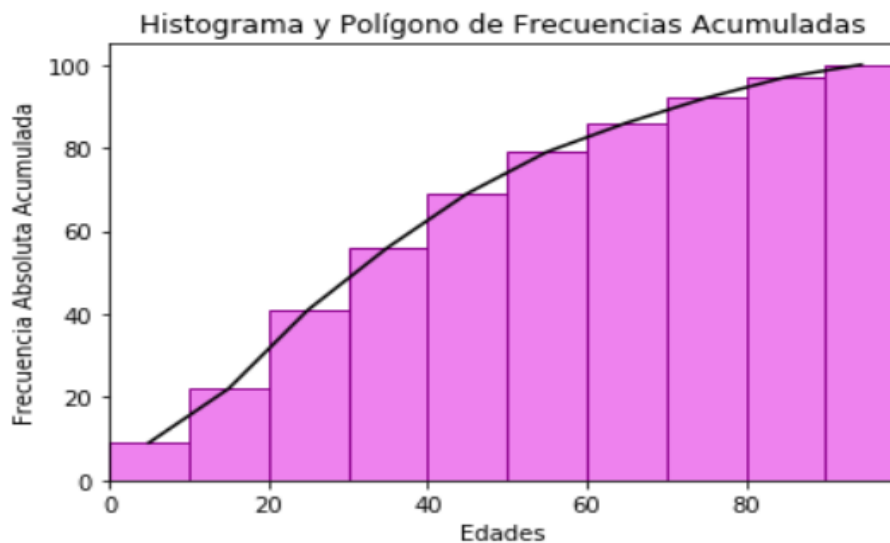
Histograma de frecuencia relativa

Un histograma de frecuencia relativa es semejante a una gráfica de barras, pero se usa para graficar cantidades en lugar de datos cualitativos. Primero, dividimos el intervalo de las mediciones más pequeñas a las más

grandes en subintervalos o clases de igual longitud. Si se ponen en columna los puntos de cada subintervalo y se traza una barra sobre cada una de las columnas, se habrá creado un histograma de frecuencia o un histograma de frecuencia relativa, dependiendo de la escala del eje vertical (Mendehall, Beaver, Beaver. 2010 Pág. 24).

Polígono de frecuencias

Es la representación gráfica que se realizan trazando los puntos que representan las frecuencias y uniéndolos mediante segmentos, es decir, es una "herramienta gráfica que se emplea a partir de un histograma de frecuencia. Para ello, se unen con una línea los distintos puntos medios de las columnas del histograma, sin dejar espacio entre una y otra, logrando así una forma geométrica o polígono" (Concepto.de. En línea).



Vale señalar que los polígonos de frecuencias se utilizan para graficar o resaltar "distintas distribuciones conjuntas o bien, una clasificación cruzada de una variable cuantitativa continua, junto con otra variable cualitativa o cuantitativa discreta, todo dentro de un mismo gráfico" (Concepto.de. 2021).

Criterios, representación y análisis de gráficas

Tablas de contingencia

La tabla de contingencia se utiliza para probar la hipótesis de que los métodos de clasificación renglón-columna son independientes. Si se rechaza la hipótesis entonces se concluye que existe alguna interacción entre los dos criterios de clasificación.

Para ello, observe el siguiente ejemplo:

Tras una elección entre dos candidatos, una encuesta de salida registró el sexo y el voto de 100 electores seleccionados de manera aleatoria y los datos se tabularon de la siguiente manera:

	Candidato A	Candidato B	Todo
Hombre	26	22	48
Mujer	39	13	52
Todo	65	35	100

Esta tabla de contingencia cuenta las respuestas según sexo y voto. El conteo en la intersección de la fila h y la columna A se denota como nhA , y representa el número de observaciones que muestra esa combinación de niveles. Por ejemplo, nhA muestra el número de encuestados masculinos que votaron por el Candidato A.

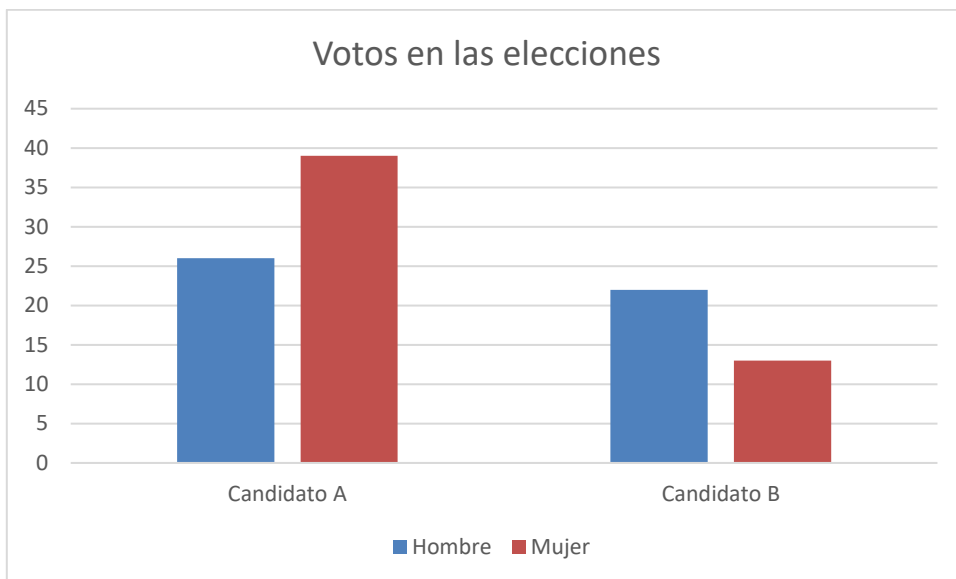
La tabla también incluye los totales marginales para cada nivel de las variables; por ejemplo, los totales marginales para las columnas muestran que 65 encuestados votaron por el Candidato A.

“Las tablas de contingencia más simples son tablas de dos factores que cuentan las respuestas según dos variables” (Minitab. 2019. En línea).

Existen diversos gráficos utilizados para la presentación de tablas de contingencia, a continuación, se presenta el gráfico de barras.

Una gráfica de barras puede orientarse verticalmente u horizontalmente y éstos representan valores numéricos, generalmente usando una hoja de cálculo. Las gráficas de barras, también denominadas diagramas de barras, o diagrama de columnas son una manera de representar frecuencias; las frecuencias están asociadas con categorías. La gráfica de barras sirve para comparar y tener una representación gráfica de la diferencia de frecuencias o de intensidad de la característica numérica de interés.

Retomando el ejemplo sobre las votaciones, la representación gráfica sería:



Actividades de aprendizaje: Revisa el siguiente enlace <https://www.superprof.es/apuntes/escolar/matematicas/estadistica/descriptiva/histograma.html> y explica qué es un histograma y cómo se conforma:

Analiza los siguientes videos:

<https://youtu.be/CuKr7Gzohbl>

<https://youtu.be/ZAJJB7gbiBs>

Con lo aprendido, construye un polígono de frecuencias y un histograma a partir de los siguientes datos:

Edad	Personas
[0, 10)	9
[10, 20)	13
[20, 30)	19
[30, 40)	15
[40, 50)	13

[50,60)	10
[60,70)	7
[70,80)	6
[80,90)	5
[90,100]	3
Total:	100

Autoevaluación

**Colegio de Bachilleres de Chiapas
Instrumento de autoevaluación del bloque 2**

Lista de cotejo para evaluar el aprendizaje de la descripción gráfica de un conjunto de datos

Nombre de la asignatura: Probabilidad y Estadística I	Bloque 2 :
Alumno:	Parcial:
Grupo:	Fecha de aplicación:

No.	Características del producto a evaluar	Ponderación: Sí =2 No =0	
1	Reconozco que es un dato.		
2	Recuerdo el concepto de frecuencia absoluta.		
3	Conozco el concepto de frecuencia relativa.		
4	Puedo calcular la frecuencia relativa acumulada.		
5	Conozco cómo calcular un intervalo		
6	Puedo construir un histograma de frecuencias.		
7	Puedo construir un polígono de frecuencias.		
	Puntos totales:		

Orientaciones para el estudio

“Aprender a aprender” es potenciar la capacidad para ser protagonistas de nuestro propio aprendizaje, tomando conciencia de nuestras

necesidades y recursos personales, así como identificando aquellas oportunidades disponibles que nos ofrece el proceso de formación profesional. Es aprender a organizar la actividad de estudio, planificar y optimizar el tiempo, adecuar las expectativas y las metas, enfrentar los obstáculos y contratiempos, manteniendo la seguridad y la confianza en sí mismo. Pero lo más importante, es aprender a disfrutar del conocimiento interactuando con él y construyéndolo en interacción.

Recursos Didácticos

Hoja electrónica de Excel
 Hojas milimétricas
 Pizarra.

Colegio de Bachilleres de Chiapas Evaluación diagnóstica Lista de cotejo

Nombre de la asignatura: Probabilidad y Estadística I		Bloque 2: Descripción gráfica de un conjunto de datos	
Alumno:		Parcial: Primero	
Grupo:		Fecha de aplicación:	
No.	Características del producto a evaluar	Ponderación: Sí =2 No =0	
1	Sabe calcular el rango en un conjunto de datos.		
2	Establece el número de clases.		
3	Sabe qué es intervalo abierto.		
4	Conoce qué es un intervalo cerrado.		
5	Puede leer la expresión matemática para establecer un intervalo.		
6	Sabe contabilizar las frecuencias en un grupo de datos.		
7	Sabe construir un histograma.		

Evaluación sumativa

Desarrolla lo que se solicita.

Los siguientes datos representan las estaturas, en metros, de 45 estudiantes de bachillerato elegidos al azar:

1.58 1.52 1.60 1.63 1.53 1.65 1.72 1.58 1.59 1.58 1.56 1.69
1.66 1.73 1.81 1.77 1.76 1.61 1.54 1.68 1.62 1.61 1.70 1.54
1.68 1.75 1.88 1.77 1.81 1.79 1.73 1.63 1.60 1.85 1.84 1.70
1.80 1.63 1.80 1.70 1.74 1.75 1.64 1.65 1.75

- Utilizando la regla de Sturges determina el número de intervalos necesarios para resumir estos datos.
- Obtener el rango.
- Calcular la amplitud para cada intervalo.
- Ubicar cada intervalo en una recta numérica asignando su forma de representación (semi-cerrados por la izquierda). Puedes utilizar el siguiente segmento rectilíneo.
- Construir las diferentes distribuciones de frecuencias (Absolutas, relativas, acumuladas y relativas acumuladas).

Intervalos	Frecuencias Absolutas	Frecuencias Relativas	Frecuencias Acumuladas	Frecuencias Relativas Acumuladas

Bloque 3: Medidas estadísticas



Introducción al bloque de aprendizaje

En este bloque 3 demuestra generalizaciones de comportamientos, así como sesgos en diferentes poblaciones objeto de cualquier ámbito presente en su entorno, por medio del uso creativo de las medidas de tendencia central y dispersión.

Objetivos específicos de aprendizaje del bloque

- Examina de manera crítica información resumida por medio de las medidas de tendencia central obtenidas en diferentes contextos.
- Emplea comparaciones entre la dispersión de datos que le permitan estimar su comportamiento y los datos que arrojan las medidas estadísticas para la toma de consciente de decisiones.
- Usa límites estadísticos para el análisis crítico y reflexivo de situaciones problema presente en su contexto.

Competencias Genéricas/Disciplinarias

Competencias Genéricas	Competencias Disciplinarias
5. Desarrolla innovaciones y propone soluciones a problemas a partir de métodos establecidos. CG5.1 Sigue instrucciones y procedimientos de manera reflexiva, comprendiendo cómo cada uno de sus pasos contribuye al alcance de un objetivo. CG5.5 Sintetiza evidencias	CDEM 1. Construye e interpreta modelos matemáticos mediante la aplicación de procedimientos aritméticos, algebraicos, geométricos y variacionales, para la comprensión y análisis de situaciones reales, hipotéticas o formales. CDEM 2 Formula y resuelve problemas matemáticos aplicando

<p>obtenidas mediante la experimentación para producir conclusiones y formular nuevas preguntas.</p> <p>8 Participa y colabora de manera efectiva en equipos diversos.</p> <p>CG8.3 Asume una actitud constructiva, congruente con los conocimientos y habilidades con los que cuenta dentro de distintos equipos de trabajo.</p>	<p>diferentes enfoques.</p> <p>CDEM 3 Explica e interpreta los resultados obtenidos mediante procedimientos matemáticos y los contrasta con modelos establecidos o situaciones reales.</p> <p>CDEM 4 Argumenta la solución obtenida de un problema, con métodos numéricos, gráficos, analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las tecnologías de la información y la comunicación.</p>
---	---

Interdisciplinariedad/Ejes transversales

Interdisciplinariedad	Ejes transversales
<ul style="list-style-type: none"> • Geografía. • Estructura Socioeconómica de México. • Se retomarán las asignaturas que en cada plantel se impartan en 5o. Semestre, tanto del Componente de Formación Propedéutica como el de Formación para el Trabajo. 	<ul style="list-style-type: none"> • Eje transversal social. • Eje transversal ambiental. • Eje transversal de salud. • Eje transversal de habilidades lectoras.

Conceptos clave

Media
Mediana
Moda
Sesgo
Rango
Desviación media
Varianza
Desviación estándar

Cuartiles
Deciles
Percentiles.

Bloque 3. Medidas estadísticas

Actividad Inicial. Evaluación diagnóstica

1. Escribe con tus propias palabras qué entiendes por medidas de tendencia central

2. Al término del primer semestre, Eduardo obtuvo las siguientes calificaciones:

Matemáticas 1	7
Química 1	8
Ética 1	9
Metodología de la investigación	10
Taller de Lectura y Redacción	9
Inglés 1	8
Informática 1	9

Calcula el promedio de Eduardo al concluir el primer Semestre

3. Las edades de un grupo de amigos son: 18, 17, 19, 20, 17, 20, 21. ¿Cuál de las siguientes opciones indica la mediana de las edades?

En la siguiente tabla se describen los ingresos semanales de 10 empleados de una pequeña empresa de transporte. (Contestar pregunta 4 y 5):

Cantidad de empleados	Salario Semanal
-----------------------	-----------------

2	\$800
4	\$900
3	\$1,200
1	\$1,500

4. ¿Cuál de las siguientes opciones indica el promedio de los ingresos semanales de los 10 empleados?

- a) 1130 b) 1100 c) 1030 d) 1200

5. ¿Cuál es la moda de la tabla anterior?

- a) 800 b) 900 c) 1200 d) 1500

Medidas de tendencia central para datos simples y agrupados

Las medidas de tendencia central o medidas de centralización nos indican donde se concentran la mayoría de los datos de una muestra y nos sirven como punto de referencia para poder ubicar o comparar cualquier dato de la muestra analizada, entre las más comunes están la media aritmética, mediana y moda.

Media aritmética

Conocida también como promedio o media es la medida de centralización más utilizada. El valor de la media aritmética se obtiene al sumar todos los datos y dividir el resultado entre el número total de datos. Se utiliza el símbolo \bar{x} (equis barra) para denotar la media aritmética de una muestra el símbolo μ (letra griega mu) para denotar la media aritmética de una población. La fórmula para calcular la media aritmética para **datos no agrupados (simples)** es:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{N} \quad \bar{x} = \frac{\text{suma total de todos los datos}}{\text{total de datos}}$$

Donde:

x_1, x_2, x_3 son los datos de la muestra

N es el total de datos

Ejemplo: Calcular la media aritmética de los siguientes datos no agrupados

Datos	Procedimiento	Resultado
3, 5, 8, 7, 9	$\bar{x} = \frac{3 + 5 + 8 + 7 + 9}{5} = \frac{32}{5} = 6.4$ $x_1 = 3, x_2 = 5, x_3 = 8, x_4 = 7, x_5 = 9$ $N = 5$	$\bar{x} = 6.4$
4, 7, 3, 8, 5, 9, 6	$\bar{x} = \frac{4 + 7 + 3 + 8 + 5 + 9 + 6}{7} = \frac{42}{7} = 6$	$\bar{x} = 6$
12, 15, 11, 17, 14, 13, 18, 20	$\bar{x} = \frac{12 + 15 + 11 + 17 + 14 + 13 + 18 + 20}{8}$ $= \frac{120}{8} = 15$	$\bar{x} = 15$

Media aritmética de datos agrupados

Cuando agrupamos los datos en una tabla de distribución de frecuencias tenemos una nueva variable, "la frecuencia" que son las veces que se repite un dato, por lo que la fórmula para calcular la media aritmética de **datos agrupados** es:

$$\bar{x} = \frac{(x_1 * f_1) + (x_2 * f_2) + (x_3 * f_3) \dots}{N}$$

$$\bar{x} = \sum_{i=1}^n \frac{x_i * f_i}{N}$$

Donde:

$x_i = \text{Marca de clase}$

$f_i = \text{Frecuencia}$

$N = \text{Total de datos de la muestra}$

Ejemplos: Calcular la media aritmética de los datos agrupados siguientes

Datos			Procedimiento	Resultado
2, 3, 4, 3, 4, 5, 6, 4, 5, 6			$\bar{x} = \frac{(x_1 * f_1) + (x_2 * f_2) + (x_3 * f_3)}{N}$ $\bar{x} = \frac{(1.5 * 1) + (3.5 * 5) + (5.5 * 4)}{10}$ $\bar{x} = \frac{(1.5) + (17.5) + (22)}{10}$ $\bar{x} = \frac{41}{10} = 4.1$	$\bar{x} = 4.1$
Interv- lo	Mar- ca de clase	Frecuen- cia		
1-2	1.5	1		
3-4	3.5	5		
5-6	5.5	4		
TOTAL		10		
Donde:				
$x_1 = 1.5 \quad f_1 = 1$				
$x_2 = 3.5 \quad f_2 = 5$				
$x_3 = 5.5 \quad f_3 = 4$				
1, 2, 2, 3, 4, 8, 9, 5, 6, 9, 5, 7			$\bar{x} = \frac{(x_1 * f_1) + (x_2 * f_2) + (x_3 * f_3) + \dots}{N}$	
Interv- lo	Mar- ca de clase	Frecuen- cia		

0-2	1	3	$\bar{x} = \frac{(1 * 3) + (4 * 4) + (7 * 3) + (10 * 2)}{12}$ $\bar{x} = \frac{(3) + (16) + (21) + (20)}{12}$ $\bar{x} = \frac{60}{12} = 5$	$\bar{x} = 5$
3-5	4	4		
6-8	7	3		
9-11	10	2		
TOTAL		12		
Donde: $x_1 = 1 \quad f_1 = 3$ $x_2 = 4 \quad f_2 = 4$ $x_3 = 7 \quad f_3 = 3$ $x_4 = 10. \quad f_4 = 2$				

Mediana

Es otra medida de tendencia central o centralización que se utiliza habitualmente y es el valor que se encuentra justo en el centro de un conjunto de datos numéricos ordenados de menor a mayor. La mediana divide al conjunto de datos en dos partes iguales. Para obtener el lugar o la posición donde buscar la mediana en un conjunto de "n" observaciones se utiliza lo siguiente:

$$\text{Posición de la Mediana} = \frac{(n + 1)}{2}$$

Para **datos no agrupados (simples)** en los que el número de elementos es par, la mediana está dada por el promedio de los dos datos centrales, pudiéndose obtener un valor no dado en la muestra.

Ejemplos:

Datos	Orden de los datos	Resultado
2, 3, 1, 4, 5	1 2 3 4 5	<i>Mediana = 3</i>
4, 7, 5, 9, 11, 6, 10	4 5 6 7 9 10 11	<i>Mediana = 7</i>
2, 4, 8, 6, 12, 10	2, 4, 6, 8 , 10, 12 En este caso dos valores quedan en el centro que son 6 y 8. Se suman los dos valores centrales y se dividen entre 2. $\text{Mediana} = \frac{6 + 8}{2} = \frac{14}{2} = 7$	<i>Mediana = 7</i>

Cálculo de la mediana para datos agrupados

Para obtener la mediana de **datos agrupados** se puede utilizar cualquiera de las siguientes fórmulas.

$$M_e = L_i + \frac{c \left(\frac{n}{2} - F_{i-1} \right)}{f_i}$$

$$M_e = L_i + \frac{c(0.5 - H_{i-1})}{h_i}$$

Donde:

L_i =Límite inferior del intervalo que contiene a la mediana.

F_{i-1} = Frecuencia acumulada en la clase anterior $i - \text{ésima}$.

f_i = Frecuencia en la clase que contiene a la mediana.

H_{i-1} =Frecuencia relativa acumulada en la clase anterior $i - \text{ésima}$.

h_i = Frecuencia relativa en la clase que contiene a la mediana.

c = Tamaño del intervalo de la clase.

Ejemplo: Se realizó una encuesta en la que se preguntaba la cantidad aproximada de horas que las personas pasaban frente al celular. Calcular la mediana de los resultados de la encuesta.

Horas	Frecuencia	
0-2	6	Primero se va a encontrar cada parte de la fórmula por separado y luego se juntará todo en la fórmula para obtener el resultado de forma sencilla y ordenada.
2-4	34	
4-6	9	
6-8	1	
Total	50	
Se inicia con buscar el intervalo de la mediana. $n/2$	$n/2$ $= 50/2$	Con este resultado se encuentra el intervalo, la mediana está en el intervalo que tenga la frecuencia absoluta mayor al resultado de $n/2$, entonces, ahora se procede a encontrar las

	= 25	frecuencias acumuladas. Se marca con un "*" el intervalo de la mediana.
--	------	--

Horas	Frecuencia f_i	Frecuencia Acumulada F_i	Ahora que se encontró el intervalo de la mediana se puede conocer f_i que sería la frecuencia en la clase que contiene a la mediana que es 34, F_{i-1} que sería la frecuencia acumulada del primer intervalo que sería 6. También podemos conocer L_i que es límite inferior del intervalo que es 2.
0-2	6	6	
2-4	34	*40	
4-6	9	49	
6-8	1	50	

Lo último que hace falta es el valor de c , que es el tamaño del intervalo de la clase.

$$c = 4 - 2$$

$$c = 2$$

En el caso de la fórmula 2 se necesitan los datos de $H_{i-1} = 6/50 = 0.12$, que es la frecuencia relativa acumulada anterior que se obtiene a partir del cociente entre la frecuencia acumulada del valor determinado entre el número total de datos y $h_i = 34/50 = 0.68$ que es la frecuencia relativa en la clase que contiene a la mediana y se obtiene a partir del cociente entre la frecuencia del valor determinado entre el número total de datos.

Horas	Frecuencia	Frecuencia Acumulada	Frecuencia relativa h_i	Frecuencia acumulada H_i
-------	------------	----------------------	---------------------------	----------------------------

0-2	6	6	0.12	0.12
2-4	34	*40	0.68	0.8
4-6	9	49	0.18	0.98
6-8	1	50	0.02	1

Se calcula la mediana usando cualquiera de las dos fórmulas.

$$M_e = L_i + \frac{c \left(\frac{n}{2} - F_{i-1} \right)}{f_i}$$

$$M_e = 2 + \frac{2 \left(\frac{50}{2} - 6 \right)}{34}$$

$$M_e = 2 + \frac{2(19)}{34}$$

$$M_e = 2 + \frac{38}{34}$$

$$M_e = 2 + 1.11$$

$$M_e = 3.11$$

$$M_e = L_i + \frac{c(0.5 - H_{i-1})}{h_i}$$

$$M_e = 2 + \frac{2(0.5 - 0.12)}{0.68}$$

$$M_e = 2 + \frac{2(0.38)}{0.68}$$

$$M_e = 2 + \frac{0.76}{0.68}$$

$$M_e = 2 + 1.11$$

$$M_e = 3.11$$

Moda

La moda es una medida de tendencia central que describe “lo más frecuente” o “lo que más se repite” en un conjunto de datos que pueden

ser numéricos o cualitativos. La moda puede no existir (si ningún dato se repite) y si existe, entonces puede ser unimodal, bimodal o multimodal.

Ejemplos:

Datos	Desarrollo / Resultado
1. Calcular la moda de los siguientes datos: 1, 3, 4, 5, 3, 6, 3, 8, 2, 3, 7	El dato que más se repite es el 3 Moda=3
2. Calcular la moda de los siguientes datos: 1, 2, 3, 4, 5, 6, 7, 8, 9	Ningún dato se repite. No hay moda.
3. 1, 1, 2, 2, 3, 3, 4, 4, 5, 5	Cada dato se repite dos veces No hay moda.
4. Calcular la moda de los siguientes datos: 10, 20, 30, 30, 40, 40, 50, 60, 70	Los datos 30 y 40 son moda. Bimodal.

Moda para datos agrupados

En una distribución de frecuencias de datos agrupados, la moda es la marca del intervalo de clase que contiene la mayor frecuencia.

La moda para **datos agrupados** se puede calcular mediante la fórmula:

$$M_o = L_i + \left(\frac{d_1}{d_1 + d_2} \right) i$$

Donde:

M_o es la moda.

L_i es el límite de la frontera inferior de la clase modal (la clase modal es la clase con la más alta frecuencia en la distribución).

d_1 es la frecuencia de la clase modal menos la frecuencia de la clase

que se encuentra inmediatamente antes de ella.

d_2 es la frecuencia de la clase modal menos la frecuencia de la clase que se encuentra inmediatamente después de ella.

i es la amplitud de la clase modal (intervalo).

Ejemplo:

Calcular la moda de los siguientes datos que se recopilaron al preguntar a un grupo de estudiantes cuántas horas duermen al día.

Primero, se encuentra el intervalo de la moda, con "*" se marca el intervalo de la moda [6-8].

El límite inferior de la clase modal $L_i=6$

Horas de sueño	Frecuencia
2-4	2
4-6	7
6-8	* 21
8-10	5

Se determina la amplitud de la clase modal del intervalo i

$$i = 8 - 6$$

$$i = 2$$

Ahora se determinan los valores d_1 y d_2

$$d_1 = f_i - f_{i-1}$$

$$d_2 = f_i - f_{i+1}$$

$$d_1 = 21 - 7$$

$$d_2 = 21 - 5$$

$$d_1 = 14$$

$$d_2 = 16$$

Con la información obtenida se aplica la fórmula para determinar la moda para datos agrupados.

$$M_o = L_i + \left(\frac{d_1}{d_1 + d_2} \right) i$$

$$Mo = 6 + \left(\frac{14}{14 + 16}\right) 2$$

$$Mo = 6 + \left(\frac{14}{30}\right) 2$$

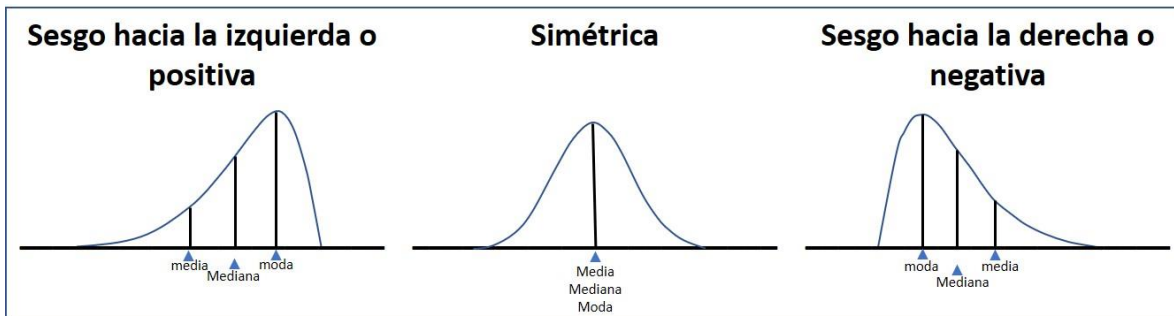
$$Mo = 6 + \left(\frac{14}{30}\right) 2$$

$$Mo = 6 + 0.93$$

$$Mo = 6.93$$

Sesgo

Con frecuencia una distribución no es simétrica alrededor de ningún valor, pero en lugar de ello se tiene que los datos están aglomerados o distribuidos hacia los extremos. Si hay pocos datos distribuidos hacia el extremo derecho se dice que la distribución es sesgada a la derecha, mientras que, si hay pocos datos distribuidos hacia la izquierda, se dice que la distribución es sesgada hacia la izquierda. Las medidas que describen esta asimetría se denominan coeficiente de sesgo, o simplemente sesgo.



El sesgo evalúa el grado de distorsión o inclinación que adopta la distribución de los datos respecto a su valor promedio tomado como centro de gravedad. El coeficiente de asimetría de Pearson es:

$$A_s = \frac{3(\bar{x} - Me)}{S}$$

Donde:

\bar{x} = media aritmética o promedio

Me = Mediana

$S =$ Desviación estándar

Grado de asimetría	
Simetría perfecta.	Cero. El promedio es igual a la mediana.
Sesgo positivo.	Positivo. Promedio mayor que la mediana.
Sesgo negativo.	Negativo. Promedio menor que la mediana.

Para conocer más sobre las medidas de tendencia central puedes ver el video de YouTube a través del siguiente link:

<https://www.youtube.com/watch?v=Bro2lbFIYPQ>

Actividad 1 Medidas de tendencia central

1. Recopila las edades de los compañeros de tu grupo y escribe los datos en el siguiente espacio:

2. Calcula la media aritmética, mediana y moda de los datos anteriores. Realiza las operaciones en el espacio siguiente:

Media aritmética	Mediana	Moda

3. Agrupa los datos (los valores de las edades del ejercicio 1) en una tabla de distribución de frecuencias. Utiliza el siguiente espacio:

4. A partir de la tabla de distribución de frecuencias del ejercicio anterior, calcula nuevamente la media aritmética, la mediana y la moda.

Media Aritmética	Mediana	Moda

Medidas de dispersión para datos simples y agrupados

Las medidas de dispersión también conocidas como medidas de descentralización, estas medidas o puntos de referencia obtenidos, se utilizan para analizar los valores numéricos de las muestras y saber que tan dispersos están con respecto a su media, entre las más comunes están el rango, **la desviación media, varianza y desviación estándar.**

Rango

Es considerado como la medida de dispersión más simple para el análisis de datos. No ofrece mucha información sobre la variabilidad de los datos por estar basada solo en los valores extremos, razón por la cual debe ser usada como complemento de otras medidas de dispersión, indica la distancia que existe entre el dato menor y el dato mayor en un conjunto de datos. Para el cálculo del rango se utiliza la siguiente fórmula:

$$\text{Rango} = \text{Dato mayor} - \text{Dato menor}$$

Rango para datos agrupados

$$\text{Rango} = \text{Límite superior de la última clase} \\ - \text{Límite inferior de la primera clase}$$

Ejemplo:

Datos	Desarrollo	Resultado
Según el servicio meteorológico nacional en las últimas 24 horas se registraron las siguientes temperaturas en la costa chiapaneca: 29°, 33°, 37°, 41°	$\text{Rango} = 41^{\circ}\text{C} - 29^{\circ}\text{C} = 12^{\circ}\text{C}$	La temperatura varió 12°C ese día.

La desviación media

La desviación de un dato es el valor que indica “que tan alejado” está de la media aritmética del grupo al que pertenece. La desviación es positiva si el dato es mayor que la media y la desviación es negativa si el dato es menor que la media aritmética; la suma de todas las desviaciones es igual a cero. Al calcular el valor absoluto de la desviación de cada dato se ob-

tiene su desviación absoluta. La desviación media DM es el promedio de las desviaciones absolutas de los datos, su fórmula es:

$$DM = \frac{\sum_{i=1}^N |dato_i - \bar{x}|}{N}$$

Donde:

$DM =$ Desviación media

$\bar{x} =$ media aritmética

$dato_i =$ los datos del ejercicio

$|dato_i - \bar{x}| =$ Valor absoluto

$N =$ Total de datos de la muestra

Ejemplo:

Pasos	Desarrollo
1. Calcular la media aritmética de los datos del ejemplo anterior. 29°C, 33°C, 37°C, 41°C	$\bar{x} = \frac{\sum_{i=1}^n x_i}{N}$ $\bar{x} = \frac{29 + 33 + 37 + 41}{4} = \frac{140}{4} = 35$
2. Calcular la desviación media.	$DM = \frac{ 29 - 35 + 33 - 35 + 37 - 35 + 41 - 35 }{4}$

	$DM = \frac{6 + 2 + 2 + 6}{4} = \frac{16}{4} = 4$
3. Resultado:	Cada una de las 4 temperaturas monitoreadas varió en promedio 4°C de la temperatura media (35°C).

Varianza

Es el promedio de las desviaciones elevada al cuadrado de cada dato de la muestra con respecto a la media aritmética. Sirve para determinar qué tan alejados se encuentran los datos de la muestra con respecto a su media aritmética. Se representa por el símbolo S_2 y su fórmula es:

$$S_2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

Donde:

$S_2 = \text{Varianza}$

$x_i = \text{Datos de la muestra}$

$\bar{x} = \text{Media aritmética de la muestra}$

$N = \text{Total de datos de la muestra}$

Desviación estándar

Es la medida de dispersión más utilizada sobre todo en las áreas de producción, mientras más pequeño sea su valor más preciso son los datos, es decir, es muy baja su variabilidad. Nos permite, además, comparar dos muestras diferentes en las que se analiza una misma variable que tiene igual media aritmética

El símbolo S representa la desviación estándar de una muestra y equivale a la raíz cuadrada de la varianza. La fórmula para calcular la desviación estándar es:

$$S = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

$S =$ Desviación estándar

$x_i =$ Datos de la muestra

$\bar{x} =$ Media aritmética de la muestra

$N =$ Total de datos de la muestra

Ejemplo. Calcular en la varianza y desviación estándar con base en los siguientes datos: 29, 33, 37, 41

<p>1. Se calcula la media aritmética de los datos:</p> <p>29°, 33°, 37°, 41°</p>	$\bar{x} = \frac{\sum_{i=1}^n x_i}{N}$ $\bar{x} = \frac{29 + 33 + 37 + 41}{4} = \frac{140}{4} = 35$
<p>2. Se calcula la varianza:</p>	$S_2 = \frac{(29 - 35)^2 + (33 - 35)^2 + (37 - 35)^2 + (41 - 35)^2}{4}$ $S_2 = \frac{(-6)^2 + (-2)^2 + (2)^2 + (6)^2}{4}$

	$S_2 = \frac{36 + 4 + 4 + 36}{4} = \frac{80}{4} = 20$ $S_2 = 20$
<p>3. Se calcula la desviación estándar que corresponde a la raíz cuadrada de la varianza:</p>	$S = \sqrt{S_2}$ $S = \sqrt{20}$ $S = 4.47$

Para conocer más sobre las medidas de dispersión puedes ver el video de YouTube a través del siguiente link:

<https://www.youtube.com/watch?v=VjCeoPLmbhl>

También, puedes acceder a la plataforma de Khan Academy donde puedes aprender en línea sobre los temas de Probabilidad y Estadística I por medio de videos y ejercicios en línea de forma divertida a través del siguiente link:

<https://es.khanacademy.org/>

Actividad 2 Medidas de dispersión

En cada ejercicio:

- Realiza la distribución de frecuencia de los datos.

- Localiza la clase modal y calcula la media aritmética.
- Calcula el rango y la desviación estándar de los datos no agrupados.
- Analiza las relaciones entre dos o más medidas de dispersión de la población para determinar su comportamiento.

Desarrolla tu actividad en tu libreta de apuntes o en un documento de Word.

1. Un profesor se interesa en estudiar el tiempo que le lleva a los alumnos leer un artículo de Taller de Lectura y Redacción, por lo que recoge datos acerca del número de segundos que tarda cada alumno en terminar de leer el artículo, los datos son los siguientes:

80	90	93	93	84
81	81	86	87	88
95	97	95	82	85
94	95	95	84	83
92	90	87	84	85

2. El docente de matemáticas presenta las calificaciones para un determinado grupo de 30 estudiantes del curso de cálculo diferencial.
Las calificaciones son las siguientes:

9	8	9	7	9
8	6	7	8	7
7	8	6	6	6
6	9	9	9	8
7	7	7	7	7
9	6	7	8	6

Límites estadísticos

“Son aquellas entidades numéricas utilizadas para señalar la posición que ocupa un dato determinado, en relación con el resto de datos numéricos, permitiendo así conocer otros puntos propios de la distribución de datos, que no son inherentes a los valores centrales” (Escobar, Guillermo. 2019). Entre estos se encuentran:

Cuartiles, deciles y percentiles

Las medidas de posición, también llamadas cuantiles, son aquellas que permiten calcular valores en la distribución de los datos y que la dividen en partes iguales, de tal forma, que los intervalos generados por los cuantiles contienen el mismo número de datos. Los cuantiles más usados son los **cuartiles, deciles y percentiles**. Cuando se tienen datos agrupados en intervalos, estas medidas se consideran en cierta forma como una extensión de la mediana.

Cuartiles

Son tres valores Q_1 , Q_2 , y Q_3 , que dividen a los datos en cuatro partes iguales. Existen tres cuartiles y cada una de las partes representa el 25%.

El cálculo de los cuartiles se realiza mediante el siguiente procedimiento:

1. Ordenar los datos de forma ascendente.
2. Calcular la posición i con la ecuación $i = \left\lfloor \left(\frac{k}{4}\right) n \right\rfloor$. Donde k es el número del cuartil ($k=1, 2, 3$) y n es el número total de datos.
3. Si i no es un número entero se debe redondear al entero siguiente y el valor que ocupa esta posición será el cuartil requerido. Si i es un número entero, el cuartil es el promedio de i e $i + 1$.

Ejemplo: se encuestó a 7 estudiantes sobre el número de horas a la semana que dedican al repaso de los temas vistos en clase. Los resultados son los siguientes: 3, 5, 2, 7, 6, 4, 9 horas.

1. Ordenar los datos.	2, 3, 4, 5, 6, 7, 9
2. Para el cuartil Q_1 , la posición i sería:	$i = \left\lfloor \left(\frac{1}{4}\right) 7 \right\rfloor = 1.75$

3. Dado que i no es un entero, se redondea al entero siguiente.	$i = 1.75$ Al redondear $i=2$
En este caso, el cuartil Q_1 corresponde al valor ubicado en la posición 2, el cual es 3 horas.	Su interpretación significa que el 25% de los estudiantes dedican máximo 3 horas semanales para el repaso a los temas vistos en clase.

Deciles

Los deciles D_k son valores que fraccionan la distribución de los datos en diez partes iguales. En la distribución se presentan nueve deciles: el D_1 acumula el 10% de los datos, el D_2 deja el 20%, y así sucesivamente hasta el D_9 que acumula el 90% de los datos.

Para el cálculo de los deciles se usa un procedimiento similar al de los cuartiles.

1. Ordenar los datos de forma ascendente.
2. Calcular la posición i con la ecuación $i = \left\lfloor \left(\frac{k}{10}\right)n \right\rfloor$. Donde k es el número del decil ($k=1, 2, 3, 4, 5, 6, 7, 8, 9$) y n es el número total de datos.
3. Si i no es un número entero se debe redondear al entero siguiente y el valor que ocupa esta posición será el decil requerido. Si i es un número entero, el decil es el promedio de i e $i + 1$.

Para datos agrupados en intervalos: $D_k = l_i + \left[\frac{k\left(\frac{n}{10}\right) - N_{i-1}}{n_i} \right] * C$

Ejemplo: Con base en la siguiente tabla de distribución de frecuencias determinar el decil tres.

Intervalo	x_i	f_i	F_i
[10, 15)	12.5	3	3
[15, 20)	17.5	5	8

[20, 25)	22.5	7	15
[25, 30)	27.5	4	19
[30,35)	32.5	2	21

1. Buscamos el intervalo donde se encuentra el tercer decil, multiplicando 3 por N=21 y dividiendo entre 10. $\rightarrow \frac{21(3)}{10} = \frac{63}{10} = 6.3$
2. Buscamos en la columna de frecuencias acumuladas (**Fi**) el intervalo que tiene a 6.3
La clase de D_3 es [15,20)
3. Se aplica la fórmula para el cálculo de deciles para datos agrupados, extrayendo los siguientes datos:

$$L_i = 15$$

$$F_{i-1} = 3$$

$$f_i = 5$$

$$a_i = 5$$

$$D_k = l_i + \left[\frac{k \left(\frac{n}{10} \right) - N_{i-1}}{n_i} \right] * C = 15 + \left[\frac{6.3 - 3}{5} \right] * 5 = 18.3$$

$$D_3 = 18.3$$

Percentiles

Los percentiles (P_k) son valores que fraccionan la distribución de los datos en cien partes iguales. En la distribución se presentan 99 percentiles: el primer percentil P_1 acumula el 1% del conjunto de datos, el percentil P_2 deja el 2%, y de forma similar los demás percentiles hasta llegar al percentil P_{99} que acumula el 99% de los datos. Para el cálculo de los percentiles se usa un procedimiento similar al empleado para los cuartiles y deciles.

1. Ordenar los datos de forma ascendente.
2. Calcular la posición i con la ecuación $i = \left\lfloor \left(\frac{k}{100} \right) n \right\rfloor$. Donde k es el número del percentil ($k=1, 2, 3, 4, 5, \dots, 98, 99$) y n es el número total de datos.

3. Si i no es un número entero se debe redondear al entero siguiente y el valor que ocupa esta posición será el percentil requerido. Si i es un número entero, el percentil es el promedio de i e $i + 1$.
4. Para datos agrupados en intervalos: $P_k = l_i + \left[\frac{k\left(\frac{n}{100}\right) - N_{i-1}}{n_i} \right] * C$

Al analizar los cuartiles, deciles y percentiles se pueden deducir las siguientes relaciones:

$$Q_2 = D_5 = P_{50} = Me$$

$$Q_1 = P_{25}$$

$$Q_3 = P_{75}$$

$$D_1 = P_{10}$$

$$D_2 = P_{20}$$

$$D_3 = P_{30}$$

$$D_4 = P_{40}$$

$$D_6 = P_{60}$$

$$D_7 = P_{70}$$

$$D_8 = P_{80}$$

$$D_9 = P_{90}$$

Actividad 3 Límites estadísticos

Resuelve los siguientes ejercicios:

En un examen de ingreso a la universidad se aplicó una prueba sobre habilidad matemática. Los resultados obtenidos por los estudiantes que aspiraban a estudiar las carreras de Ingeniería civil y Arquitectura, fueron los siguientes:

Aspirantes a Ingeniería Civil

3.0	3.7	2.5	2.2	3.5	3.0	3.5	2.8	3.0	3.0
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

2.7	3.8	3.8	2.2	2.5	3.8	2,2	4.0	2.3	2.8
3.3	2.8	2.7	3.5	2.2	2.7	3.7	3.8	2.5	3.9
2.2	3.2	2.3	2.3	2.0	3.8	2.7	3.3	2.5	3.2
2.7	2.2	3.2	2.5	3.7	2.3	3.0	3.8	2.5	3.7
3.5	2.3	2.0	2.5	3.0	2.8	3.0	3.0	2.2	2,3
3.8	2.2	3.0	2.0	2.3	3.7	3.9	3.0	2.4	3.8
2.7	2.3	2.3	2.7	2.0	2,2	2.3	2.1	2.7	3.2

Aspirantes a Arquitectura

3.8	3.5	2.7	2.2	2.3	2.2	2.4	3.4	3.6	2.8
2.7	3.7	3.3	3.8	2.0	2.0	2,6	3.5	2.0	2.2
3.7	2.0	3.0	3.8	2.3	2.5	3.2	3.4	2.7	4.0
2.7	3.3	2.2	2.2	2.3	3.8	2.9	3.1	2.4	3.8
2.0	2.7	3.2	3.8	3.3	2.5	2.9	3.9	2.2	3.3
2.5	2.3	2.7	3.5	3.5	2.2	3.4	3.1	2.5	3.3

1. Calcular los cuartiles y deciles de cada uno de los datos presentados en las tablas.
2. Interpretar los resultados en cada una de las tablas.
3. Comparar los cuartiles entre los ejercicios de las tablas.

Realizar las operaciones requeridas en la libreta de apuntes o en un documento en Word.

Autoevaluación

Lee detenidamente las preguntas y responde colocando una X en el nivel de avance que consideras que has logrado a lo largo del bloque I.

Interpretación del nivel de avance

10% → Excelente, logré el aprendizaje de manera independiente.

8% → Bueno, requerí apoyo para construir mi aprendizaje.

6% → Regular, fue difícil el proceso de aprendizaje y lo logré parcialmente.

5% o menos → Insuficiente, no logré el aprendizaje.

Contenidos	Nivel de avance			
	10%	8%	6%	5% o menos
1. Identificas correctamente las medidas de tendencia central.				
2. Identificas los datos simples y agrupados.				
3. Reconoces el concepto de media, mediana y moda.				
4. Aplicas las medidas de tendencia central en datos simples.				
5. Aplicas las medidas de tendencia central en datos agrupados.				
6. Identificas las medidas de dispersión.				
7. Identificas los conceptos de rango, desviación media, varianza y desviación estándar.				
8. Aplicas las medidas de dispersión en datos simples y agrupados.				
9. Identificas los límites estadísticos.				
10. Aplicas correctamente los conceptos de cuartiles, deciles y percentiles en problemas cotidianos.				
Total				

Orientaciones para el estudio

Hoy en día, los jóvenes de la Educación Media Superior (EMS) transitan hacia la vida adulta, interactúan en un mundo que evoluciona de la sociedad del conocimiento hacia la sociedad del aprendizaje y la innovación (Joseph Stiglitz, 2014; Ken Robinson, 2015; Richard Gerver, 2013; y Marc Prensky, 2015; entre otros); procesan enormes cantidades de información a gran velocidad y comprenden y utilizan, de manera simultánea, la tecnología que forma parte de su entorno cotidiano y es relevante para sus intereses.

Por lo anterior, en la Educación Media Superior debe superarse la desconexión existente entre el currículo, la escuela y los alumnos, ya que la misma puede producir la desvinculación educativa de éstos, lo cual, incluso puede derivar en problemas educativos como los bajos resultados, la reprobación y el abandono escolar. Para ello, en primer lugar, hay que entender que los jóvenes poseen distintos perfiles y habilidades (no son un grupo homogéneo) que requieren potenciar para desarrollar el pensamiento analítico, crítico, reflexivo, sintético y creativo, en oposición al esquema que apunte sólo a la memorización; esto implica superar, asimismo, los esquemas de evaluación que dejan rezagados a muchos alumnos y que no miden el desarrollo gradual de los aprendizajes y competencias para responder con éxito al dinamismo actual que los jóvenes requieren enfrentar para superar los retos del presente y del futuro. En segundo lugar, se requiere un currículo pertinente y dinámico, en lugar del vigente que es segmentado y limitado por campo disciplinar, que se centre en la juventud y su aprendizaje, y que ponga énfasis en que ellos son los propios arquitectos de sus aprendizajes.

La escuela, en consecuencia, requiere transformarse de fondo para lograr incorporar en el aula y en la práctica docente las nuevas formas en que los jóvenes aprenden, y lo seguirán haciendo (Gerver, 2013; Prensky, 2013); de no hacerlo, quedará cada día más relegada de la realidad.

Recursos didácticos

Computadora
Pizarrón
Hoja de cálculo.

Evaluación

Tema integrador del Bloque 3:

El docente formará equipos de trabajo de tres personas, donde el trabajo de cada grupo será encuestar a un número de personas de su comunidad, en el caso de la escuela un grupo, por ejemplo, o hacerlo en la hora de receso entrevistando a personas de manera aleatoria.

El tema a tratar queda a criterio de cada equipo, por ejemplo:

- El tiempo que dedican los estudiantes a repasar apuntes en una semana, los deportes favoritos de los alumnos del plantel, o bien analizar alguna información relevante de estudio en la que se pueda aplicar las medidas estadísticas vistas en este bloque.

El equipo tiene que recopilar y organizar los datos y aplicar las medidas estadísticas de acuerdo a los datos obtenidos e interpretarlos.

Las medidas de tendencia central

- Media, mediana y moda.

Medidas de dispersión

- Rango
- Desviación media
- Varianza
- Desviación estándar.

Limites Estadísticos

- Cuartiles, deciles y percentiles.

Se hace una rúbrica donde se califica: la encuesta, la tabla de distribución de datos, la presentación del trabajo con proyector, y la forma de explicar la interpretación de los resultados.

Bloque 4: Comportamiento de dos variables

Introducción



En muchas ocasiones es importante analizar la relación entre dos variables cuantitativas de un grupo de entes o sujetos, los objetivos de dichos análisis suelen ser:

- Determinar si dos variables están correlacionadas, es decir, si los valores de una variable tienden a ser más altos o más bajos para valores más altos o más bajos de la otra variable.
- Poder predecir el valor de una variable dado un valor determinado de la otra variable.
- Valorar el nivel de concordancia entre los valores de las dos variables.

Objetivos específicos de aprendizaje del bloque

Al término del presente bloque, el alumno tendrá la capacidad de analizar si dos variables están relacionadas entre sí, el grado de correlación, la ecuación que mejor la representa y la gráfica correspondiente, proponer hipótesis y realizar pruebas de bondad de ajuste.

Competencias Genéricas/Disciplinares

<p>Competencias Genéricas:</p> <ul style="list-style-type: none">✓ 5.4 Construye hipótesis, diseña y aplica modelos para probar su validez.✓ 5.5 Sintetiza evidencias obtenidas mediante la experimentación para producir conclusiones y formular nuevas preguntas.✓ 9.5 Actúa de manera propositiva frente a fenómenos de la sociedad y se mantiene informado.	<p>Competencias Disciplinares:</p> <ul style="list-style-type: none">✓ CDEM 2 Formula y resuelve problemas matemáticos aplicando diferentes enfoques✓ CDEM 4 Argumenta la solución obtenida de un problema con métodos numéricos, gráficos y analíticos o variacionales, mediante el lenguaje verbal, matemático y el uso de las Tecnologías de Información y la Comunicación.✓ CDEM 8 Interpreta tablas, gráficas, mapas, diagramas y textos con símbolos matemáticos y científicos.
---	---

Ejes transversales

Eje transversal Social: se sugiere retomar temas relacionados con la educación financiera, moral y cívica, para la paz (derechos humanos), equidad de género, interculturalidad, lenguaje no sexista, vialidad entre otros.

Eje transversal Ambiental: se recomienda abordar temas referentes a la naturaleza, uso de recursos naturales, desarrollo sustentable, reciclaje entre otros.

Eje transversal de Salud: se sugiere abordar temas relacionados con la educación sexual integral y reproductiva, cuidado de la salud, prevención y consumo de sustancias tóxicas entre otros.

Eje transversal de Habilidades Lectoras: se recomienda retomar temas relacionados con la lectura, comprensión lectora, lecto-escritura y lectura de textos comunitarios o en lenguas nativas entre otros.

Conceptos clave

Variable dependiente, variable independiente, diagrama de dispersión, regresión lineal simple, ecuación de regresión, ecuación estimada de regresión, método de los mínimos cuadrados, coeficiente de determinación, coeficiente de correlación.

Actividad Inicial. Evaluación diagnóstica

Responde correctamente las siguientes preguntas planteadas:

1. ¿A qué se le llama variable independiente?

- a) Una variable continua
- b) Una variable compleja
- c) Una variable cuyo valor se obtiene a través de una regla de relación
- d) Una variable cuyo valor no se asocia a una regla de relación

2. ¿Qué es un diagrama de dispersión?

- a) Una gráfica de puntos
- b) Una gráfica bidimensional que ubica los puntos relativos a un experimento
- c) Una muestra de puntos
- d) Una serie de puntos que siguen una regla determinada

3. ¿Para qué sirve el método de mínimos cuadrados?

- a) Para obtener el cuadro más pequeño
- b) Para hallar la recta de regresión lineal
- c) Para localizar la mínima diferencia entre un punto y la recta
- d) Para hacer la diferencia entre los puntos lo más pequeña posible

4. ¿Para qué sirve la recta de regresión?

- a) Para predecir nuevos valores de la variable independiente
- b) Para proponer nuevos valores de la variable independiente
- c) Para revisar su correlación
- d) Para construir un diagrama de dispersión

5. ¿Qué valores toma el coeficiente de correlación?

- a) 1
- b) -1
- c) 0
- d) -1 a +1

Actividad 2. El alumno revisará los siguientes enlaces:

<https://youtu.be/31OwPATjrc> Diagrama de dispersión

<https://youtu.be/cH0AeVt1BQU> Diagrama de dispersión



Diagrama de dispersión

El diagrama de dispersión se usa comúnmente para mostrar como dos variables se relacionan entre sí. Ambas variables se representan como un punto en el plano cartesiano y de acuerdo a la relación que exista entre ellas, definimos su tipo de correlación.

“Se pueden interpretar varios tipos de correlación a través de los patrones mostrados en los diagramas de dispersión. Estos son: **positivo** (los valores

aumentan juntos), **negativo** (un valor disminuye a medida que el otro aumenta), **nulo** (sin correlación), **lineal**, **exponencial** y **en forma de U**. La fuerza de la correlación puede determinarse por la proximidad de los puntos entre sí en el gráfico. Los puntos que terminan muy lejos del conjunto general de puntos se conocen como valores atípicos." (datavizcatalogue. En línea)

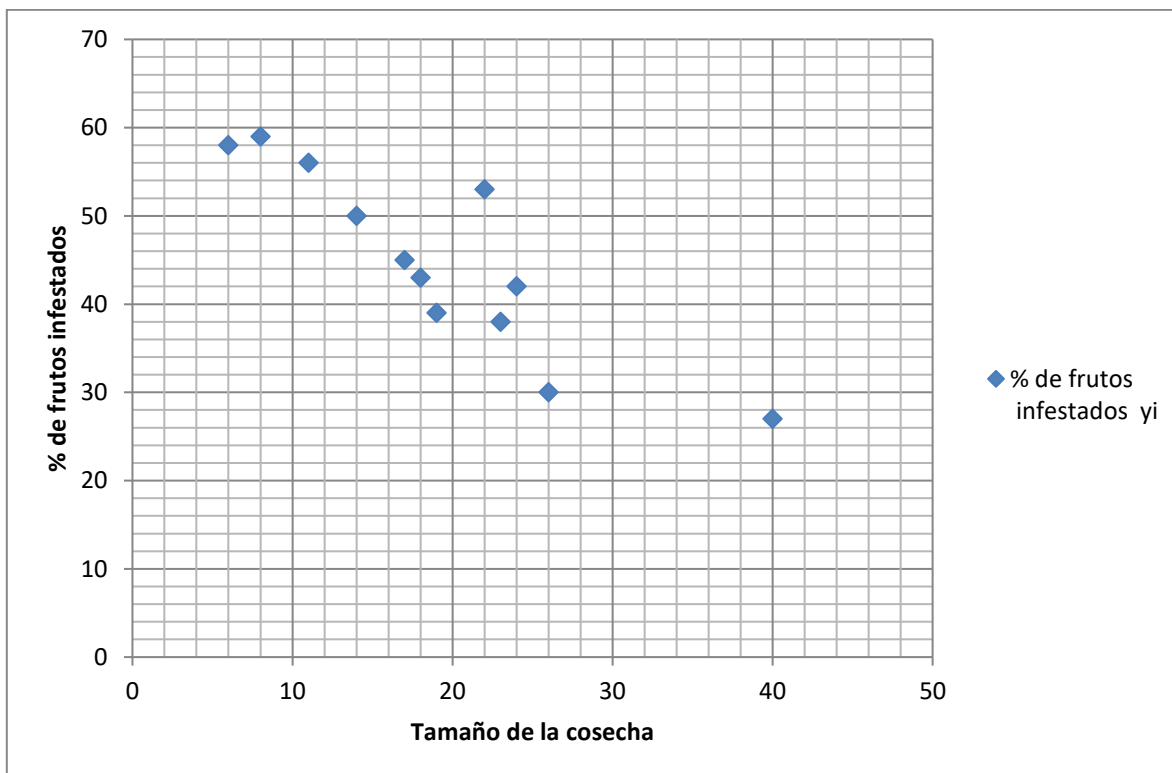
Actividad 3. Elabora un diagrama de dispersión en una hoja milimétrica con los siguientes datos:

Doce árboles fueron sometidos a una aspersion en el cáliz de la flor con un químico llamado arseniato de plomo, seguido de cinco aspersiones globales consistentes en tres libras de arseniato de plomo y un cuarto de aceite de pescado por cada 100 galones. Los datos del % de frutas infestadas y tamaño de la cosecha se muestran enseguida, se quiere conocer la relación que existe entre las dos variables.

Árbol	Tamaño de la cosecha * 100 frutos	% de frutos infestados
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
12	40	27

Actividad 4. Elabora un diagrama de dispersión en Excel.

La respuesta del alumno debe ser como la siguiente:



Regresión lineal

¿Qué es el análisis de regresión?

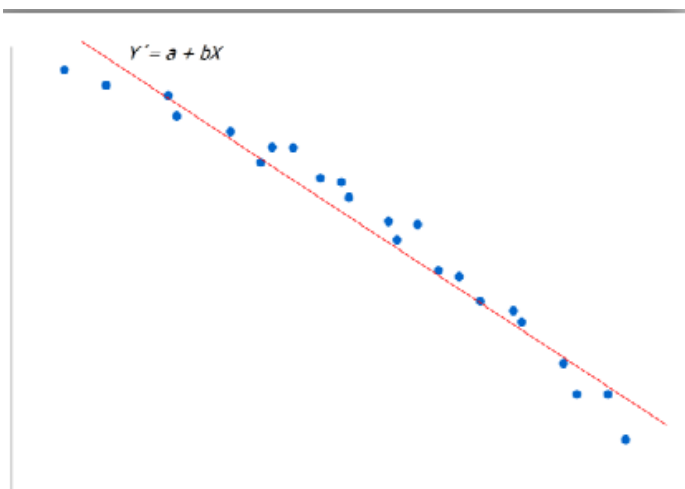
Es la técnica empleada para realizar la predicción del valor de la variable dependiente Y , con base en un valor seleccionado de la variable independiente X .

Nota: La palabra *regresión* la utilizó por primera vez Sir Francis Galton en 1877 en su estudio de los factores hereditarios. Descubrió que las estaturas de los descendientes de padres altos, tendían a una regresión (es decir a volver o retornar) hacia la estatura promedio de la población.

Al observar el diagrama de dispersión se puede considerar que una línea recta parece describir mejor la ubicación promedio de los puntos, por lo que se determinará mediante una ecuación matemática correspondiente a una línea recta.

¿Qué es la ecuación de regresión?

Es una expresión matemática que define la relación entre dos variables, llamada también recta de regresión. Se pueden trazar manualmente varias



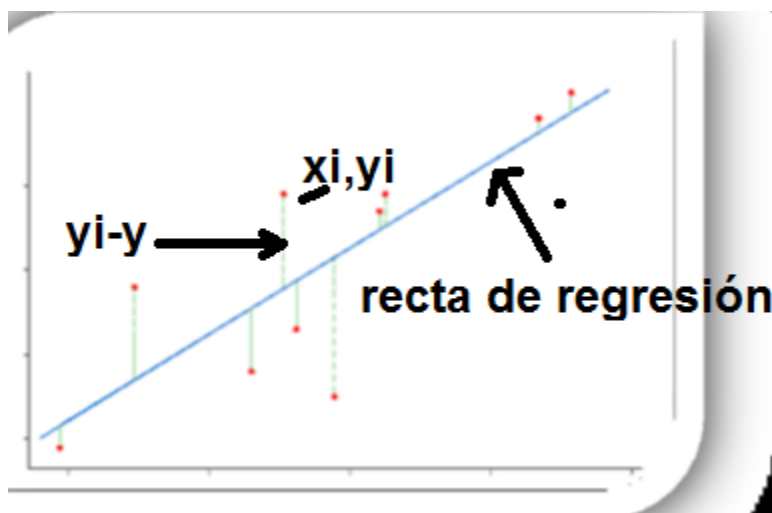
rectas que pasen aproximadamente cerca de todos los puntos, pero el concepto de "cerca" se debería al juicio de cada persona que realiza el ajuste; para evitar esta subjetividad y elegir la recta que mejor se ajuste a los puntos, utilizaremos el método de mínimos cuadrados.

¿Qué es el método de mínimos cuadrados?

Es una técnica empleada para llegar a la ecuación de regresión, minimizando la suma de los cuadrados de las distancias o desviaciones verticales entre los valores Y verdaderos (de los puntos) y los valores pronosticados o estimados de "y" de la recta.

Fundamentación Matemática:

Se tienen los puntos x_i, y_i



$y = mx + b$ recta de regresión

Como x_i, y_i son puntos correspondientes a los datos, se consideran constantes, por otra parte m y b corresponden a la recta que tiene una posición variante, por lo que se consideran variables.

Entonces la diferencia entre la recta y un punto la llamaremos d_i , como el método es de mínimos cuadrados, es decir, la distancia d_i siempre debe ser positiva, para ello la elevamos al cuadrado, definiendo la suma de las diferencias al cuadrado:

$$\sum_{i=1}^n d_i^2$$

Nótese que la diferencia entre el punto d_i y la recta se define como $d_i = y_i - (mx_i + b)$.

La cual es una función de m y b ; de tal modo que la función queda expresada así:

$$f(m, b) = \sum_{i=1}^n (y_i - mx_i - b)^2$$

Ecuación 1

Esta expresión debe ser minimizada, es decir, debe tener la mínima distancia entre los puntos y la recta de regresión; para ello, es necesario derivar e igualar a cero; como depende de dos variables, se deriva parcialmente, las derivadas son respecto a las variables m y b .

Derivando respecto a m :

$$\begin{aligned} \frac{\partial f(m,b)}{\partial m} &= \sum_{i=1}^n \frac{\partial}{\partial m} [(y_i - mx_i - b)^2] \\ \frac{\partial f(m,b)}{\partial m} &= \sum_{i=1}^n 2 \left[(y_i - mx_i - b) \frac{\partial}{\partial m} (y_i - mx_i - b) \right] \\ \frac{\partial f(m,b)}{\partial m} &= \sum_{i=1}^n 2 \left[(y_i - mx_i - b) \frac{\partial}{\partial m} (y_i) - \frac{\partial}{\partial m} mx_i - \frac{\partial}{\partial m} b \right] \end{aligned}$$

Como x_i , y_i son constantes y al derivar parcialmente la segunda variable permanece constante, la derivada de una constante es cero, por tanto:

$$\frac{\partial}{\partial m} (y_i) = 0, \quad \frac{\partial}{\partial m} b = 0, \quad \frac{\partial}{\partial m} mx_i = x_i$$

Por tanto, la derivada parcial queda:

$$\frac{\partial f(m,b)}{\partial m} = \sum_{i=1}^n 2[(y_i - mx_i - b)(-x_i)]$$

Multiplicando por x_i tenemos:

$$\sum_{i=1}^n (-2x_i y_i + 2mx_i^2 + 2bx_i) = 0$$

Aplicando la sumatoria:

$$-2 \sum_{i=1}^n x_i y_i + 2 \sum_{i=1}^n x_i^2 m + 2 \sum_{i=1}^n x_i b = 0$$

Dividiendo entre 2 y transponiendo el primer término:

$$\left(\sum_{i=1}^n x_i^2 \right) m + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n x_i y_i$$

Ecuación 2

Derivando la ecuación #1 respecto de b , tenemos:

$$\frac{\partial f(m,b)}{\partial b} = \sum_{i=1}^n \frac{\partial}{\partial b} [(y_i - mx_i - b)^2]$$

$$\frac{\partial f(m,b)}{\partial b} = \sum_{i=1}^n 2(y_i - mx_i - b) \frac{\partial}{\partial b} (y_i - mx_i - b)$$

$$\frac{\partial f(m,b)}{\partial b} = \sum_{i=1}^n 2(y_i - mx_i - b) \left[\frac{\partial}{\partial b} (y_i) - \frac{\partial}{\partial b} mx_i - \frac{\partial}{\partial b} b \right]$$

Nuevamente x_i , y_i son constantes y m permanece constante al derivar parcialmente, así tenemos:

$$\frac{\partial}{\partial b} (y_i) = 0, \frac{\partial}{\partial b} mx_i = 0, \frac{\partial}{\partial b} b = -1$$

Sustituyendo:

$$\frac{\partial f(m,b)}{\partial b} = \sum_{i=1}^n 2(y_i - mx_i - b)(-1)$$

Multiplicando por -1 e igualando a cero para MINIMIZAR:

$$-2 \sum_{i=1}^n y_i + 2 \sum_{i=1}^n mx_i + 2 \sum_{i=1}^n b = 0$$

Dividiendo entre 2 y transponiendo:

$$\sum_{i=1}^n mx_i + \sum_{i=1}^n b = \sum_{i=1}^n y_i$$

Recordando que la sumatoria

$$\sum_{i=1}^n b = nb$$

Tenemos:

$$\sum_{i=1}^n mx_i + nb = \sum_{i=1}^n y_i$$

Despejando b

$$nb = \sum_{i=1}^n y_i - \sum_{i=1}^n mx_i$$

Transponiendo n

$$b = \frac{\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i}{n}$$

Sustituyendo el valor de b en la ecuación # 2

$$\left(\sum_{i=1}^n xi^2\right)m + \left(\sum_{i=1}^n xi\right)\frac{\sum_{i=1}^n yi - m \sum_{i=1}^n xi}{n} = \sum_{i=1}^n xiyi$$

Multiplicando los términos de la fracción:

$$\left(\sum_{i=1}^n xi^2\right)m + \frac{\sum_{i=1}^n xi \sum_{i=1}^n yi}{n} - \frac{m \sum_{i=1}^n xi \sum_{i=1}^n xi}{n} = \sum_{i=1}^n xiyi$$

Factorizando en m

$$m \left[\sum_{i=1}^n xi^2 - \frac{\sum_{i=1}^n xi \sum_{i=1}^n xi}{n} \right] + \frac{\sum_{i=1}^n xi \sum_{i=1}^n yi}{n} = \sum_{i=1}^n xiyi$$

Transponiendo el segundo término:

$$m \left[\sum_{i=1}^n xi^2 - \frac{\sum_{i=1}^n xi \sum_{i=1}^n xi}{n} \right] = \sum_{i=1}^n xiyi - \frac{\sum_{i=1}^n xi \sum_{i=1}^n yi}{n}$$

Despejando m

$$m = \frac{\sum_{i=1}^n xiyi - \frac{\sum_{i=1}^n xi \sum_{i=1}^n yi}{n}}{\sum_{i=1}^n xi^2 - \frac{\sum_{i=1}^n xi \sum_{i=1}^n xi}{n}}$$

Multiplicando numerador y denominador por n

$$m = \frac{n \sum_{i=1}^n xiyi - \sum_{i=1}^n xi \sum_{i=1}^n yi}{n \sum_{i=1}^n xi^2 - (\sum_{i=1}^n xi)^2}$$

Resumen de las fórmulas:

$$m = \frac{n \sum_{i=1}^n xiyi - \sum_{i=1}^n xi \sum_{i=1}^n yi}{n \sum_{i=1}^n xi^2 - (\sum_{i=1}^n xi)^2}$$

$$b = \frac{\sum_{i=1}^n yi - m \sum_{i=1}^n xi}{n}$$

Calcular primero m y después b para la recta de regresión:

$$y = mx + b$$

De lo anterior, lo que más importa son las fórmulas obtenidas en el resumen.

Continuando con el ejemplo, resolvamos el ejercicio con ayuda de Excel.

Árbol	Tamaño de la cosecha * 100 frutos x_i	% de frutos infestados y_i
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
n=12	40	27

Para este ejemplo $n=12$, la fórmula requiere la suma del producto de $x_i y_i$
 $\sum_{i=1}^n x_i y_i$

La suma de x_i $\sum_{i=1}^n x_i$, la suma de y_i $=\sum_{i=1}^n y_i$, la suma de los cuadrados de x_i $=\sum_{i=1}^n x_i^2$

La suma de x_i elevada al cuadrado $=\left(\sum_{i=1}^n x_i\right)^2$

Agreguemos en Excel las columnas faltantes:

Árbol	Tamaño de la cosecha * 100 frutos x_i	% de frutos infestados y_i	$x_i y_i$ es el producto de x_i por y_i	x_i elevado al cuadrado
1	8	59	472	64
2	6	58	348	36
3	11	56	616	121

4	22	53	1166	484
5	14	50	700	196
6	17	45	765	289
7	18	43	774	324
8	24	42	1008	576
9	19	39	741	361
10	23	38	874	529
11	26	30	780	676
12	40	27	1080	1600
Suma	228	540	9324	5256
	$\sum_{i=1}^n x_i$	$\sum_{i=1}^n y_i$	$\sum_{i=1}^n x_i y_i$	$\sum_{i=1}^n x_i^2$

Apliquemos la fórmula:
$$m = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$\sum_{i=1}^n x_i y_i = 9324$	$\sum_{i=1}^n x_i = 228$
$\sum_{i=1}^n y_i = 540$	$\sum_{i=1}^n x_i^2 = 5256$

$$m = \frac{12(9324) - 228(540)}{63072 - (228)^2}$$

$$m = \frac{111888 - 123120}{63072 - 51984}$$

$$m = \frac{-11232}{11088}$$

$$m = 1.0298701$$

Calculando b

Fórmula:

$$b = \frac{\sum_{i=1}^n yi - m \sum_{i=1}^n xi}{n}$$

Tomando los datos del recuadro anterior: $b = \frac{540 - (-1.0298701)(228)}{12}$

$$b = \frac{540 + 230.961039}{12}$$

$$b = 64.2467532$$

Por tanto, la recta de regresión es:

$$y = mx + b$$

$$y = -1.029x + 64.246$$

Solicite a los alumnos ver el siguiente video:

<https://youtu.be/LI14ftR6EVs> línea de tendencia Excel.

Cálculo del coeficiente de correlación

La finalidad de la correlación es examinar la dirección y la fuerza de la asociación entre dos variables cuantitativas. Así conoceremos la intensidad de la relación entre ellas y *si, al aumentar el valor de una variable, aumenta o disminuye el valor de la otra variable.*

Coefficiente de correlación lineal de Pearson

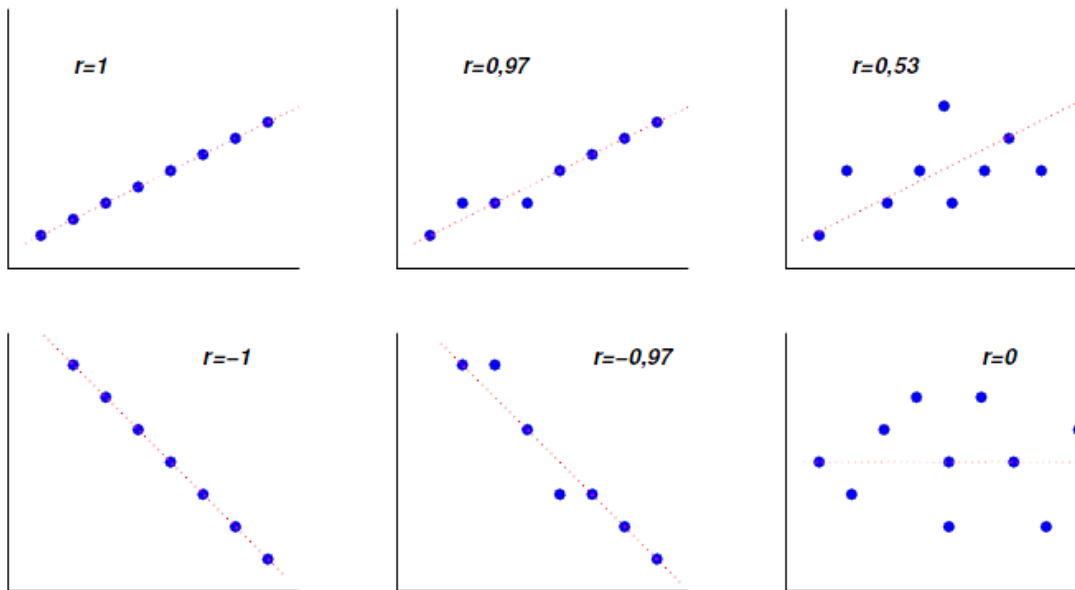
El estimador muestral más utilizado para evaluar la asociación lineal entre dos variables X e Y es el **coeficiente de correlación de Pearson (r)**. Se trata de un índice que mide si los puntos tienen tendencia a disponerse en una línea recta. Puede tomar **valores entre -1 y +1**.

Volviendo al **coeficiente de correlación lineal r**, veamos qué **propiedades** tiene:

- Carece de unidades de medida (adimensional).
- Sólo toma valores comprendidos entre [-1,1].

- Cuando $|r|$ esté **próximo a uno**, $r = +1$ (recta lineal creciente de izquierda a derecha) o $r = -1$ (recta lineal decreciente), se tiene que existe una **relación lineal muy fuerte** entre las variables.
- Cuando $r \approx 0$, puede afirmarse que **no existe relación lineal** entre ambas variables. Se dice en este caso que las variables son **in-correlacionadas**.

Para entenderlo mejor, veamos los siguientes diagramas de dispersión:



Definiciones:

La suma de las desviaciones cuadradas de "x"

$$\sum d^2x = \sum x^2 - \frac{(\sum x)^2}{n}$$

La suma de las desviaciones cuadradas de "y"

$$\sum d^2y = \sum y^2 - \frac{(\sum y)^2}{n}$$

La suma de los productos de las desviaciones de "x" y "y"

$$\sum dxdy = \sum xy - \frac{\sum x \sum y}{n}$$

El coeficiente de correlación $r = \frac{\sum dxdy}{\sqrt{\sum d^2x * \sum d^2y}}$

Árbol	Tamaño de la cosecha * 100 frutos x_i	% de frutos infestados y_i	$x_i y_i$ es el producto de x_i por y_i	x_i elevado al cuadrado	y_i^2
1	8	59	472	64	3481
2	6	58	348	36	3364
3	11	56	616	121	3136
4	22	53	1166	484	2809
5	14	50	700	196	2500
6	17	45	765	289	2025
7	18	43	774	324	1849
8	24	42	1008	576	1764
9	19	39	741	361	1521
10	23	38	874	529	1444
11	26	30	780	676	900
12	40	27	1080	1600	729
Suma	228	540	9324	5256	25522
	$\sum_{i=1}^n x_i$	$\sum_{i=1}^n y_i$	$\sum_{i=1}^n x_i y_i$	$\sum_{i=1}^n x_i^2$	$\sum_{i=1}^n y_i^2$

$$\sum dxdy = \sum xy - \frac{\sum x \sum y}{n}$$

9324 -10260 -936

$$\sum d^2x = \sum x^2 - \frac{(\sum x)^2}{n}$$

5256 -4332 924

$$\sum d^2y = \sum y^2 - \frac{(\sum y)^2}{n}$$

25522 -24300 1222

$$r = \frac{\sum dxdy}{\sqrt{\sum d^2x * \sum d^2y}} \qquad r = \frac{-936}{\sqrt{924 * 1222}}$$

$$r = -0.8808$$

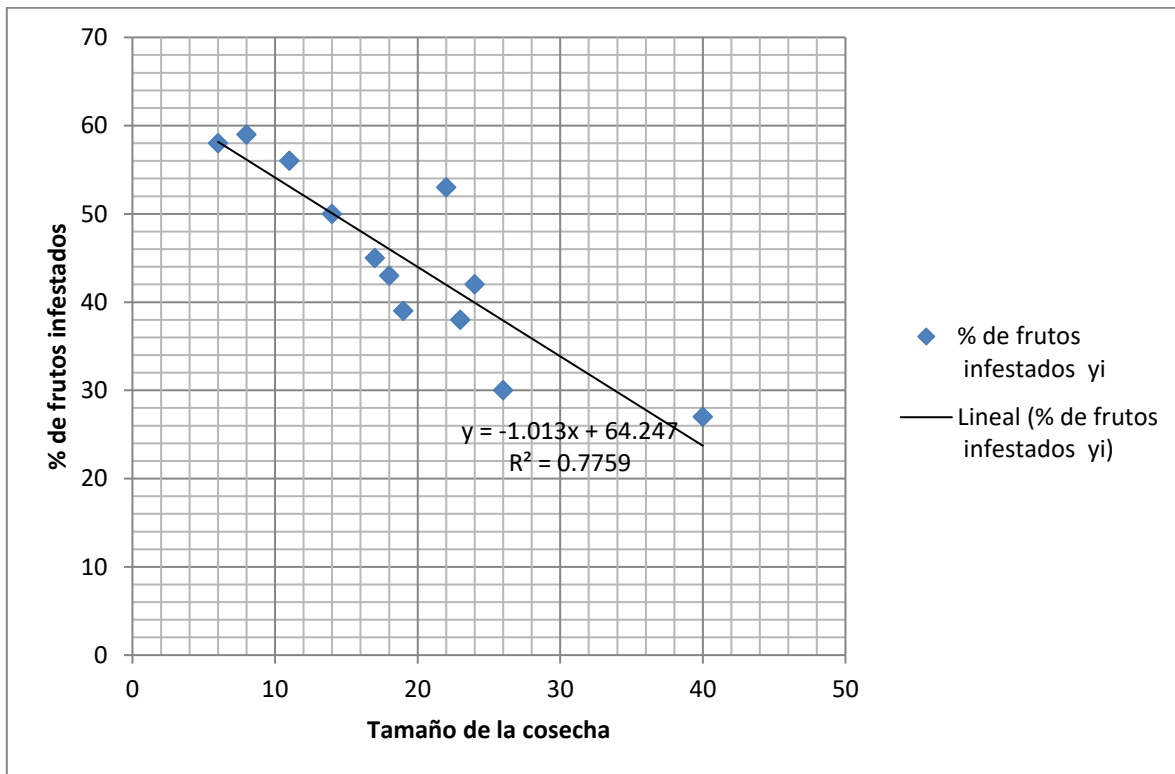
El cuadrado de r $r^2 == (-0.8808)^2 = 0.7759$

Significa que aproximadamente el 77.59 % de las variaciones en el % de frutos infestados se explican por variaciones en el tamaño de la cosecha y el otro 22.49% se deben a otros factores.

Actividad para el alumno:

En el mismo diagrama de dispersión que tienes en hoja milimétrica traza la ecuación de la recta de mínimos cuadrados.

Debe quedar como la siguiente:



Como puedes observar los datos de la gráfica fueron calculados en Excel y coinciden con los datos calculados manualmente.

Ejercicio para el alumno:

La Comisión Federal de Electricidad división Sureste, zona de distribución San Cristóbal de Las Casas, está interesada en conocer la relación que guardan las inconformidades de sus clientes con las cancelaciones de la facturación del suministro de energía eléctrica en el año 2020, datos reales obtenidos del departamento comercial.

Mes	No. de observación	Inconformidades	Cancelaciones
Enero	1	313	825
Febrero	2	326	883
Marzo	3	337	1067
Abril	4	354	862
Mayo	5	409	965
Junio	6	305	905
Julio	7	279	864
Agosto	8	210	854

Septiembre	9	263	614
Octubre	10	269	657
Noviembre	11	320	659
Diciembre	12	293	744

1. Elabora un diagrama de dispersión.
2. Aplica la regresión lineal y obtén la ecuación.
3. Elabora la gráfica con la recta de regresión y su coeficiente de correlación.

Xi cuadrada

La prueba chi-cuadrada, también llamada Ji cuadrado (X^2), de acuerdo a Ruiz Mitjana (2021. En línea) “se encuentra dentro de las pruebas pertenecientes a la estadística descriptiva, concretamente la estadística descriptiva aplicada al estudio de dos variables. Por su parte, la estadística descriptiva se centra en extraer información sobre la muestra. En cambio, la estadística inferencial extrae información sobre la población. El nombre de la prueba es propio de la distribución Chi-cuadrado de la probabilidad en la que se basa. Esta prueba **fue desarrollada en el año 1900 por Karl Pearson.**”

La prueba chi-cuadrado es una de las más conocidas y utilizadas para analizar variables nominales o cualitativas, es decir, para determinar la existencia o no de independencia entre dos variables. Que dos variables sean independientes significa que no tienen relación, y que, por lo tanto, una no depende de la otra, ni viceversa.

Así, con el estudio de la independencia se origina también un método para verificar si las frecuencias observadas en cada categoría son compatibles con la independencia entre ambas variables.

Para evaluar la independencia entre las variables se calculan los valores que indicarían la independencia absoluta, lo que se denomina “frecuencias esperadas”, **comparándolos con las frecuencias de la muestra.**

Como es habitual, la hipótesis nula (H_0) indica que ambas variables son independientes, mientras que la hipótesis alternativa (H_1) indica que las variables tienen algún grado de asociación o relación.

Así, como otras pruebas para el mismo fin, la prueba chi-cuadrado **se utiliza para ver el sentido de la correlación entre dos variables nominales o de un nivel superior.**

Analiza el siguiente video:

<https://youtu.be/gHkMGcn2MsE>

Ejemplo:

En un experimento de cruza de dos tipos de maíz se obtuvieron 773 plantas verdes, 231 doradas, 238 verdes rayadas y 59 dorado-verde-rayadas.

Se desea saber si estos resultados reafirman o contradicen la teoría Mendeliana, en la cual los cuatro grupos deben estar en las proporciones 9:3:3:1, con $\alpha=5\%$.

Solución:

Sumamos las proporciones $=9+3+3+1=16$

$$\begin{array}{l} P1=9/16 \\ P2=3/16 \\ P3=3/16 \\ P4=1/16 \end{array}$$

Hipótesis Nula $H_0=$

Hipótesis Alternativa: H_1 : Las frecuencias observadas no cumplen con estas probabilidades.

Simbología:

O_i = es la frecuencia i -ésima observada en la muestra experimental.

e_i = es la frecuencia i -ésima esperada de acuerdo a la distribución hipotética y la obtenemos multiplicando el total de las frecuencias observadas por la probabilidad teórica de ocurrencia del evento.

O_i	e_i
773	1301 $(9/16)=731.9$
231	1301 $(3/16)=243.9$

238
59

1301 (3/16)=243.9
1301 (1/16)=81.3

Suma=**1301**

Chi cuadrada calculada: x^2

$$x^2 = \sum_{i=1}^n \frac{(oi - ei)^2}{ei}$$

$$x^2 = \frac{(773 - 731.9)^2}{731.9} + \frac{(231 - 243.9)^2}{243.9} + \frac{(238 - 243.9)^2}{243.9} + \frac{(59 - 81.3)^2}{81.3}$$

$$x^2 = 2.31 + 0.68 + 0.14 + 6.12 = 9.25$$

Chi cuadrada calculada= 9.25

Este valor lo comparamos con el valor de tablas determinado por un nivel de significancia y grados de libertad:

Grados de libertad= Número de categorías -1

Grados de libertad=4-1=3

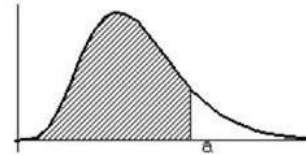
Leemos en la tabla de chi cuadrada:

La Distribución χ^2

- Cálculo de probabilidades de la χ^2

CENTILES DE LA DISTRIBUCIÓN χ^2

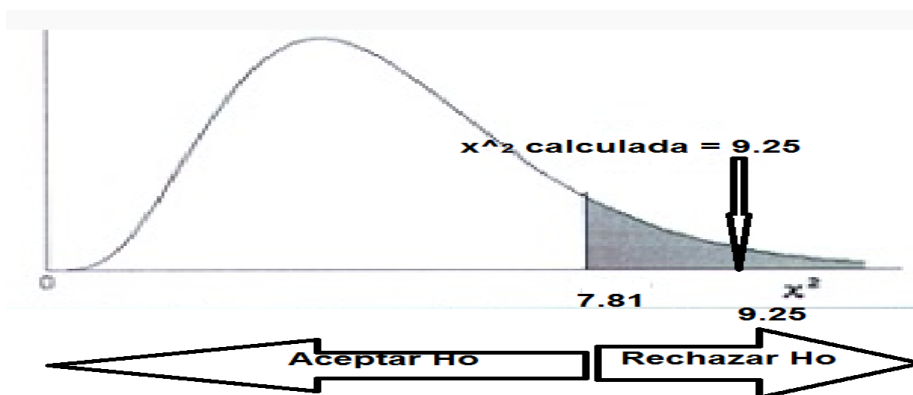
$$i = P(X \leq a)$$



n	0,995	0,99	0,975	0,95	0,9	0,75	0,5	0,25	0,05	0,025	0,01	0,005
1	7,879	6,635	5,024	3,841	2,706	1,323	0,455	0,102	0,004	0,001	0,000	0,000
2	10,597	9,210	7,378	5,991	4,605	2,773	1,386	0,575	0,103	0,051	0,020	0,010
3	12,838	11,345	9,348	7,815	6,251	4,108	2,366	1,213	0,352	0,216	0,115	0,072
4	14,860	13,277	11,143	9,488	7,779	5,385	3,357	1,923	0,711	0,484	0,297	0,207
5	16,750	15,086	12,833	11,070	9,236	6,626	4,351	2,675	1,145	0,831	0,554	0,412
6	18,548	16,812	14,449	12,592	10,645	7,841	5,348	3,455	1,635	1,237	0,872	0,676
7	20,278	18,475	16,013	14,067	12,017	9,037	6,346	4,255	2,167	1,690	1,239	0,989
8	21,955	20,090	17,535	15,507	13,362	10,219	7,344	5,071	2,733	2,180	1,646	1,344
9	23,589	21,666	19,023	16,919	14,684	11,389	8,343	5,899	3,325	2,700	2,088	1,735
10	25,188	23,209	20,483	18,307	15,987	12,549	9,342	6,737	3,940	3,247	2,558	2,156
11	26,757	24,725	21,920	19,675	17,275	13,701	10,341	7,584	4,575	3,816	3,053	2,603
12	28,300	26,217	23,337	21,026	18,549	14,845	11,340	8,438	5,226	4,404	3,571	3,074

Con 3 grados de libertad y nivel de significancia del 5% leemos en la tabla $\chi^2 = 7.815$

Dibujemos el diagrama:



Como el valor calculado es mayor que el valor leído en tablas, rechazamos la hipótesis nula, es decir, se rechaza la proporción 9:3:3:1.

Autoevaluación

El alumno se puede autoevaluar por medio de una guía de observación como la siguiente:

Colegio de Bachilleres de Chiapas
Instrumento de autoevaluación del bloque 4
Lista de cotejo para evaluar el aprendizaje del comportamiento de dos variables

Nombre de la asignatura: Probabilidad y Estadística I		Bloque 4 :	
Alumno:		Parcial:	
Grupo:		Fecha de aplicación:	
No.	Características del producto a evaluar	Ponderación: Sí=2 No =0	
1	Sé construir un diagrama de dispersión.		
2	Puedo plantear una tabla para obtener la recta de regresión.		
3	Puedo calcular la pendiente m.		
4	Ya sé calcular la ordenada al origen b.		
5	Conozco el procedimiento para calcular el coeficiente de correlación.		
6	Puedo calcular el valor de r al cuadrado.		
7	Sé interpretar el significado de r^2 .		
8	Puedo proponer la hipótesis nula y		

	alterna en la prueba de chi cuadrado.		
9	Conozco el procedimiento para leer en la tabla de chi cuadrado.		
10	Puedo llegar a una conclusión correcta aplicando el procedimiento de chi cuadrado.		
	Puntos totales		

Orientaciones para el estudio

No hay obra de arte mayor ni mecanismo más ingenioso que la simple mano del hombre. Diseñada por el gran diseñador, Isaac Newton, erudito y matemático y uno de los científicos más destacados de la historia, llegó a exclamar: "Ausentes otras pruebas, me bastaría el pulgar para convencerme de la existencia de Dios".

La mano humana, maravilla de diseño y de ingeniería, se compone de 29 huesos, 29 articulaciones, más de 100 ligamentos, 35 músculos, y un sinnúmero de nervios y arterias. Solamente para controlar el pulgar, necesitamos nueve músculos y el esfuerzo conjunto de tres nervios principales de la mano. La capacidad de la mano humana es impresionante: fuerza, flexibilidad, destreza, resistencia y control motor refinado. La punta del dedo es un instrumento sensorial dotado de una increíble capacidad para detectar, y lo hace con un grado de sensibilidad que la ingeniería humana apenas si empieza a emular con la disciplina de la robótica.

La mano de Dios es sabia: su diestra hace maravillas; qué decir del átomo o, simplemente, de un copo de nieve. Bien decía Luis Pasteur: "Un poco de ciencia aleja de Dios, pero mucha ciencia nos devuelve a él". Admiramos una computadora y un teléfono inteligente; entonces, ¡imagina el cerebro humano!

Si el tema te parece complejo o extenso, no te des por vencido, estudia con tenacidad.

Recursos didácticos

Hoja electrónica de Excel
Hojas milimétricas
Pizarra.

Evaluación del bloque: Propuesta de examen

Se examina la recaudación por rezago del impuesto predial urbano y rústico mensual durante julio-agosto 2018 y enero-junio 2019, junto con los gastos para la ejecución que implica el requerimiento de adeudos; por lo que desea obtener el modelo que se genera con la interacción entre el monto total recaudado y los gastos generados:

Mes	Gastos por requerimientos x	Recaudación por rezagos de impuesto predial y
1	2000	22360
2	2100	21685
3	1900	19961
4	2300	20778
5	1850	26337
6	2500	20200
7	2250	22500
8	1725	25139
9	2000	27332
10	2050	23702
11	2150	22451
12	2500	21580

- 1.- Elabora el diagrama de dispersión en una hoja milimétrica.
- 2.- Calcula los valores de m y b para la recta de regresión.
- 3.- En la misma gráfica de dispersión, grafica la recta de regresión.

Glosario

Cartograma: gráficos realizados sobre mapas en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan. No se recomienda su uso cuando se modifican las dimensiones del mapa.

Coefficiente de correlación: medida de la intensidad de la relación lineal entre dos variables. Coeficiente de Pearson.

Coefficiente de determinación: es la medida de bondad del ajuste de la ecuación de regresión. Se puede interpretar como la parte de la variación de la variable dependiente y que explica la ecuación de regresión.

Cuartiles: son tres valores Q_1 , Q_2 , y Q_3 , que dividen a los datos en cuatro partes iguales. Existen tres cuartiles y cada una de las partes representa el 25%.

Deciles: los deciles D_K son valores que fraccionan la distribución de los datos en diez partes iguales. En la distribución se presentan nueve deciles: el D_1 acumula el 10% de los datos, el D_2 deja el 20%, y así sucesivamente hasta el D_9 que acumula el 90% de los datos.

Desviación estándar: es la medida de dispersión más utilizada sobre todo en las áreas de producción, mientras más pequeño sea su valor más preciso son los datos, es decir, es muy baja su variabilidad. Nos permite además comparar dos muestras diferentes en las que se analiza una misma variable que tiene igual media aritmética.

Desviación media: la desviación de un dato es el valor que indica "que tan alejado" está de la media aritmética del grupo al que pertenece, la desviación es positiva si el dato es mayor que la media y la desviación es negativa si el dato es menor que la media aritmética; la suma de todas las desviaciones es igual a cero.

Diagrama de dispersión: una gráfica de datos de dos variables en la que la variable independiente está en el eje horizontal y la variable dependiente está en el eje vertical.

Documental: la recolección de datos documental es un instrumento o técnica de investigación general cuya finalidad es obtener datos e información a partir de fuentes documentales con el fin de ser utilizados dentro de los límites de una investigación en concreto.

Ecuación de regresión: es la ecuación que describe como se relaciona el valor medio o esperado de la variable dependiente con la variable independiente; en la regresión lineal simple $y = mx + b$.

Ecuación estimada de regresión: el estimado de la ecuación de regresión determinado a partir de los datos de una muestra aplicando el método de los mínimos cuadrados.

Encuesta: conjunto de preguntas tipificadas dirigidas a una muestra representativa, para averiguar estados de opinión o diversas cuestiones de hecho.

Entrevista: se utilizan para recabar información en forma verbal a través de preguntas que propone el analista.

Estadístico: función definida sobre los valores numéricos de una muestra.

Estimación: función de los valores de una muestra que se elabora para indagar el valor de un parámetro de la población de la que procede la muestra.

Experimentación: suele mostrarse como una técnica de recolección de datos, sin embargo, la experimentación no es en absoluto una forma de recolección de datos, sino más bien una estrategia de investigación.

Frecuencia acumulada: de un valor de una variable, es la que se obtiene sumando la frecuencia absoluta correspondiente a este valor, con las frecuencias absolutas de todos los valores anteriores a él. Se simboliza con f_{ac} .

Frecuencia relativa: es la proporción de elementos que pertenecen a una categoría o valor de una variable y se obtiene dividiendo su frecuencia absoluta entre el número total de elementos y se representa con el símbolo f_r . Se puede expresar en fracción, con valores decimales o en porcentajes.

Frecuencia relativa acumulada: a un valor de una variable, a la que se obtiene sumando la frecuencia relativa correspondiente a este valor, con las frecuencias relativas de todos los valores anteriores a él. Se simboliza con f_{rac} . Se puede expresar en fracción, en forma decimal o en porcentaje.

Media aritmética: conocida también como promedio o media, es la medida de centralización más utilizada. El valor de la media aritmética se obtiene al sumar todos los datos y dividir el resultado entre el número total de datos.

Mediana: es la medida de tendencia central o centralización que se utiliza habitualmente y es el valor que se encuentra justo en el centro de un conjunto de datos numéricos ordenados de menor a mayor. La mediana divide al conjunto de datos en dos partes iguales.

Método de los mínimos cuadrados: procedimiento que se usa para determinar la ecuación estimada de regresión. Se expresa $\sum(y_i - \hat{y})$.

Moda: es una medida de tendencia central que describe "lo más frecuente" o "lo que más se repite" en un conjunto de datos que pueden ser numéricos o cualitativos. La moda puede no existir (si ningún dato se repite) y si existe, entonces puede ser unimodal, bimodal o multimodal.

Muestra: subconjunto representativo de una población.

Observación: otra técnica útil para el analista en su progreso de investigación, consiste en observar a las personas cuando efectúan su trabajo.

Parámetro: función definida sobre los valores numéricos de características medibles de una población.

Percentiles: los percentiles (P_k) son valores que fraccionan la distribución de los datos en cien partes iguales. En la distribución se presentan 99 percentiles: el primer percentil P_1 acumula el 1% del conjunto de datos, el percentil P_2 deja el 2%, y de forma similar los demás percentiles hasta llegar al percentil P_{99} que acumula el 99% de los datos.

Polígono de frecuencias: es un gráfico de líneas en el cual el eje horizontal representa los datos a través de sus marcas de clase, y el eje vertical las frecuencias de cada uno de los intervalos. Para trazarlo, primero se localizan los puntos correspondientes a cada intervalo, la primera coordenada corresponde a la marca de clase y la segunda la frecuencia correspondiente. Para poder cerrar la figura, se habrá de considerar un intervalo imaginario con frecuencia cero en cada uno de los extremos de la gráfica, una vez delimitados todos los puntos, se unen de forma consecutiva con segmentos de línea recta. El polígono de frecuencias permite recuperar la idea de continuidad de la variable. El polígono puede

ser aproximado mediante una curva suavizada que suele llamarse curva de frecuencias.

Población: conjunto de individuos o elementos que cumplen ciertas propiedades y entre los cuales se desea estudiar un determinado fenómeno.

Rango: es considerado como la medida de dispersión más simple para el análisis de datos. No ofrece mucha información sobre la variabilidad de los datos por estar basada solo en los valores extremos, razón por la cual debe ser usada como complemento de otras medidas de dispersión, indica la distancia que existe entre el dato menor y el dato mayor en un conjunto de datos.

Regresión lineal simple: análisis de regresión donde interviene una variable independiente y una variable dependiente; en ella la relación entre las variables se aproxima mediante una línea recta.

Sesgo: el sesgo evalúa el grado de distorsión o inclinación que adopta la distribución de los datos respecto a su valor promedio tomado como centro de gravedad.

Técnica de muestreo probabilístico: los métodos de muestreo probabilístico buscan que todos los elementos que conforman la población tengan igual probabilidad al ser seleccionados en la muestra.

Técnica de muestreo no probabilístico: pese a que estos métodos no permiten generalización, porque todos los elementos de la población no poseen la misma probabilidad de ser seleccionados, es posible elegir las unidades muestrales bajo determinados criterios buscando mayor representatividad en la muestra.

Variables: característica observable que varía entre los diferentes individuos de una población. Las variables pueden ser cualitativas y cuantitativas.

Variables continuas: son aquellas que pueden tomar cualquier valor entre dos valores dados. Es decir, el rango contiene no solo valores enteros sino un intervalo (finito o infinito) de valores reales (esto es, que puede ser fraccionario, decimal o irracional).

Variables cualitativas: Son aquellas que representan atributos de los elementos y no permiten una representación numérica definida.

Variables cuantitativas: estas variables permiten una escala numérica y las características de los elementos son observados cuantitativamente a través de una medida y escala definida.

Variable dependiente: la variable que se predice o se explica se representa por "y".

Variables discretas: son aquellas que solo toman valores enteros con rango finito.

Variable independiente: la variable que sirve para predecir o explicar se representa por "x".

Variables nominales: son variables que no se les puede asignar un orden.

Variables ordinales: son variables que tienen forma natural de ordenación.

Varianza: es el promedio de las desviaciones elevada al cuadrado de cada dato de la muestra con respecto a la media aritmética. Sirve para determinar qué tan alejados se encuentran los datos de la muestra con respecto a su media aritmética.

Fuentes bibliográficas de consulta

Fuenlabrada, S. (2001). Probabilidad y Estadística. México: McGraw Hill.

Hurley, Aguilar Márquez, Garibay Bermúdez. Estadística. Centro de Investigación y de Estudios Avanzados. Departamento de Matemáticas. Instituto Politécnico Nacional.

Leithold, Louis. El cálculo. Oxford University Press.

Lipschutz Seymour. (2001). Probabilidad, Editorial McGraw-Hill, 2da Edición.

Mendenhall, Beaver y Beaver. (2010). Introducción a la probabilidad y estadística. Cengage Learning.

Posada Hernández Gabriel J. (2016). Elementos básicos de estadística descriptiva para el análisis de datos. (Recurso electrónico). Medellín – Funlam.

Triola, M. F. (2000). Estadística elemental. México: Pearson Educación.

Walpole, R. E., y Myers, R. H. (2012). Probabilidad y estadística. México: McGraw-Hill.

Fuentes de Internet

- <http://www.cobachsonora.edu.mx/files/semestre5-2016/probabilidadyestadistica.pdf>
- <http://www.cuentame.inegi.org.mx/monografias/informacion/chis/poblacion/educacion.aspx?tema=me&e=07>
- Datavizcatalogue.com (2021) Diagrama de dispersión. https://datavizcatalogue.com/ES/metodos/diagrama_de_dispersion.html#:~:text=Se%20pueden%20interpretar%20varios%20tipos,y%20en%20forma%20de%20U.
- Escobar, Guillermo (2019) Límites estadísticos. <https://probaesr.blogspot.com/2019/01/limites-estadisticos-son-aquell->

