



Estadística Descriptiva Básica Aplicada a la Educación Física

Autores

BORIS ANÍBAL ORTEGA ARANDA
ELZEBIR GISELA TEJEDOR DE LEÓN

ISBN: 978-9962-13-847-1



Estadística Descriptiva Básica Aplicada a la Educación Física

Autores

BORIS ANÍBAL ORTEGA ARANDA

ELZEBIR GISELA TEJEDOR DE LEÓN

© NOMBRES DE LOS AUTORES: Boris Aníbal Ortega Aranda y Elzebir Tejedor De León

© TÍTULO DEL LIBRO: Estadística descriptiva básica aplicada a la Educación Física

ISBN papel:

ISBN pdf: 978-9962-13-847-1

Depósito legal:

Reservados todos los derechos. Salvo excepción prevista por la ley, no se permite la reproducción total o parcial de esta obra, su incorporación a un sistema informático ni su transmisión en cualquier forma o por cualquier medio (electrónico, mecánico, fotocopia, grabación u otros), sin autorización previa y por escrito de los titulares del copyright. La infracción de dichos derechos conlleva sanciones legales y puede constituir un delito contra la propiedad intelectual.

Dedicatoria

A familiares, amigos y compañeros de trabajo

CONTENIDO

Presentación	vii
Introducción	ix
Capítulo 1. Conceptos introductorios básicos	1
1. Estadística.....	1
1.1. Concepto de Estadística	2
1.2. Clasificación de la Estadística	4
1.2.1. Estadística Descriptiva	4
1.2.2. Estadística Inferencial	5
1.3. Términos estadísticos básicos	6
1.3.1. Datos y series estadísticos	6
1.3.2. Universo, población y muestra	7
1.3.3. Variables	11
1.3.4. Niveles de medición	13
1.3.5. Tipos de escalas	13
Capítulo 2. Estadística e Investigación.....	17
2.1. La estadística y la investigación	17
2.2. La medición en las Ciencias Sociales	19
2.3. Naturaleza de las técnicas estadísticas utilizables en las Ciencias Sociales.....	21
Capítulo 3. Estadística Descriptiva Básica	25
3.1. Estadística Descriptiva Básica y su aplicación en el ámbito de la Educación Física.....	25
3.2. Distribuciones de Frecuencias	26
3.2.1. Concepto.....	26
3.2.2. Tipos de distribuciones de frecuencias	26

3.3 Tablas de frecuencias y su construcción.	29
3.4 Representación gráfica de distribuciones de frecuencias	32
3.4.1 Histograma	32
3.4.2. El polígono de frecuencias	34
3.5 Los cuadros y gráficos estadísticos.	38
3.6 Los gráficos	43
3.7. Las medidas de tendencia central	55
3.7.1. Medidas de tendencia central para datos simples.....	55
3.7.2. Medidas de tendencia central para datos agrupados	57
3.8 Medidas de dispersión	61
Capítulo 4. Análisis de regresión	67
4.1. Definición	67
4.2. Ecuación de regresión lineal.....	68
4.3. Diagrama de dispersión	68
4.4. Tipos de relación entre las variables	70
4.5. Principio de mínimos cuadrados.....	72
4.6. Correlación lineal	75
Capítulo 5. Introducción a la probabilidad y al muestreo	80
5.1. Conceptos básicos de probabilidad.....	80
5.2. Muestreo.....	82
5.3. Conceptos básicos en muestreo.....	84
5.4 Técnicas de muestreo	86
5.5. Selección y tamaño de muestra.....	93
REFERENCIAS	99

Presentación

En esta obra de Estadística Descriptiva básica aplicada a la Educación Física, usted encontrará la información sobre el papel que desempeña la Estadística cuando es aplicada en el ámbito de la Educación Física. Además, de ser una guía fácil para presentar y analizar diferentes tipos de datos que se pueden recolectar en actividades relacionadas con el mundo del deporte.

En las primeras páginas encontrará información general sobre conceptos estadísticos, así como también, el rol que desempeña la Estadística cuando es aplicada en el ámbito de la Educación Física. Luego, se presenta una serie de conceptos relacionados con la aplicación de la Estadística Descriptiva a la investigación en Educación Física, este segmento, ayudará a relacionar las múltiples aplicaciones de la Estadística en diversos escenarios, y el caso particularmente en el de la Educación Física.

Por último, pero no menos importante, usted encontrará una sencilla guía de problemas para resolver y poner en práctica los conceptos que se abordaron las páginas anteriores de este tratado. El presente volumen se presenta así: Introducción a la Estadística, conceptos y términos estadísticos que se utilizan en el ámbito de la Educación Física. Posteriormente, se abordan conceptos estadísticos que se aplican en Educación Física y finalmente, un conjunto de problemas prácticos que usted debe resolver y que le ayudarán a poder

explicar cómo la Estadística Descriptiva básica se relaciona con esta asignatura.

El ideal de los autores es compartir sus experiencias, la cual esta obra, representa el afán de coordinar y de presentar juntos materiales que, de otro modo, habría que buscar en ámbitos interdisciplinarios que por tradición han estado dispersos.

Se espera que los problemas sencillos, que se presentan aquí, le faciliten su iniciación a la Estadística aplicada a la Educación Física.

Atentamente, el editor.

Introducción

El uso de la estadística se ha extendido considerablemente, abarcando casi todos los campos del conocimiento humano, especialmente, cuando se quiere trabajar con datos claros e inteligibles y en consecuencia tomar decisiones en una pluralidad de disciplinas, entre las que se encuentran la Educación Física. Existe un acuerdo entre autores respecto a las funciones de la estadística y a los procedimientos que comprende, sobre todo, cuando se habla, específicamente de la estadística aplicada a esta disciplina, donde se impregna del estudio de la recolección, análisis e interpretación de datos.

En función de esto, el presente documento comprende nociones de Estadística Descriptiva (métodos de recolección, descripción, visualización y resumen de los datos, que pueden ser presentados en forma numérica o gráfica), ya que esta ciencia está ligada al método científico y participa en la recolección, organización, presentación y análisis de datos, tanto para la deducción de conclusiones, como para la toma de decisiones y cómo interviene activa y directamente desde el momento de la recolección de la información hasta la instancia de la comunicación de los resultados.

La información planteada es fundamental en la formación del futuro Licenciado en Humanidades con énfasis en Educación Física, así pues, es uno de los objetivos de este documento el familiarizar al estudiante con: los diferentes conceptos y métodos estadísticos en la investigación de la actividad física y el deporte, presentándole de forma sencilla,

problemas, que pueden resolverse utilizando aspectos básicos de Estadística Descriptiva.

Al final, se espera que el estudiante de Educación Física consiga los siguientes resultados: un alto grado de comprensión y dominio de la Estadística Aplicada a la Educación Física, la utilización de esta información para el diseño y puesta en práctica de trabajos de investigación en este campo y mejora en la elaboración de presentaciones de trabajos de investigación utilizando la Estadística.

CONCEPTOS INTRODUCTORIOS BÁSICOS

Con toda certeza se puede afirmar, que a muchas personas les desagrada escuchar el término “estadística”, pues se imaginan una serie de fórmulas imposibles de resolver, gráficos difíciles de entender y un montón de números sin sentido.

Este libro está dirigido, a ayudar al lector, a comprender conceptos básicos de la Estadística y su utilización en ámbito de la Educación Física. Presentando, además, una serie de problemas sencillos de resolver, así como también, mostrar de qué manera, esta nos ayuda a recolectar, organizar, presentar y analizar datos numéricos u observaciones.

1. Estadística.

Una característica de la sociedad moderna es: la abundante información que todos los días se genera y la mayoría de las veces, esta se presenta en una forma que la persona debe interpretar y evaluar (Díaz-Levicoy et al. 2017). Al realizar estos procesos, el ciudadano común, puede tomar mejores decisiones y comprender el mundo que lo rodea.

Es un hecho reciente, que la Estadística se ha incorporado, al currículo de las diferentes especialidades universitarias, ya que como señala Begg (citado en Batanero, 1999), esta disciplina es un excelente:

[...] vehículo para alcanzar las capacidades de comunicación, tratamiento de la información, resolución

de problemas, uso de ordenadores y trabajo cooperativo y en grupo, a las que se da gran importancia en los nuevos currículos. Además, la probabilidad y la estadística se pueden aplicar fácilmente, puesto que no requieren técnicas matemáticas complicadas. Sus aplicaciones, proporcionan una buena oportunidad para mostrar a los estudiantes, la utilidad de la matemática para resolver problemas reales, siempre que su enseñanza se lleve a cabo mediante una metodología heurística y activa, enfatizando la experimentación y la resolución de problemas (p.2).

Se recomienda usar de acuerdo con Martínez et al. (2015), existen muchos puntos estratégicos en el devenir del accionar humano donde se pone de manifiesto el empleo de métodos estadísticos. En forma general, todas estas expresiones nos indican que la estadística es una herramienta que ayuda a conocer la realidad. Sin embargo, tal y como lo señalan García y Matus (2016), *“también puede servir para distorsionar la verdad, si no se tiene cuidado al usar los métodos estadísticos adecuadamente y si la interpretación de los resultados lo hacen incorrectamente”* (p. 18).

1.1. Concepto de Estadística

De acuerdo con Batanero (2001), los orígenes de la Estadística son muy antiguos.

“La palabra Estadística se refiere a un sistema o método científico usado en la recolección, organización, análisis,

interpretación numérica de la información. También se puede decir que la Estadística estudia el comportamiento de los fenómenos en grupo”. (p. 3)

Desde el punto de vista, de Gil (2003), en una definición de Estadística resultarían característicos rasgos como los que se presentan a continuación:

- La Estadística: es una ciencia que tiene por objeto, el estudio de métodos y técnicas para el tratamiento de conjuntos de datos numéricos.
- Las técnicas estadísticas, permiten la descripción de conjuntos de datos y la inferencia sobre conjuntos más amplios.
- Los métodos desarrollados por la Estadística pueden aplicarse en distintos campos del saber, constituyendo un importante instrumento para el estudio científico.

La Estadística, trata de verificar la validez probabilística de los acontecimientos en la escala tiempo-espacio, también se la usa, para relacionar los eventos diarios; como la predicción del tiempo o al determinar el nivel probabilístico de las tasas de cambio de las monedas extranjeras en el mercado financiero (Bandii et al. 2007, p. 107).

Hampel (citado en Artés, 2006), establece que la Estadística *“es el arte y la ciencia de extraer información útil y relevante de un conjunto de datos empíricos”*. (p. 1).

En términos generales, se puede afirmar que, la Estadística actúa como una disciplina que sirve de puente entre los modelos matemáticos y los fenómenos reales, por lo que es de suponer que el quehacer estadístico

es muy antiguo, *“parece iniciarse con la invención del dado, hace unos 7,000 años”* (Ferreiro, 1998, p. 1). Pero antes de esto, fueron los astrágalos (pequeños huesos de animales), los que eran utilizados para jugar, adivinar, adivinar el futuro y hasta para comunicarse con los dioses.

La palabra Estadística: “se refiere a un sistema o método científico usado en la recolección, organización, análisis, interpretación numérica de la información”. También se puede decir que la Estadística *“estudia el comportamiento de los fenómenos en grupo”* (Martínez Bercandino, 2003, p.1)

1.2. Clasificación de la Estadística

El rápido desarrollo de la Estadística como ciencia y como una herramienta útil en y para la investigación, no sólo se ha quedado allí, sino que también, es indispensable para muchos aspectos de la vida profesional, todo esto ha sido impulsado por la difusión de las computadoras, el crecimiento de su potencia y rapidez de cálculo y las posibilidades de comunicación, lo que, en cierta medida, obliga a todos, a conocer aspectos básicos de esta ciencia (Batanero, 2001).

Dependiendo de la forma en que se analizan los datos, la Estadística se clasifica en:

1.2.1. Estadística Descriptiva

Es considerada como: una rama de la Estadística que trata sobre *“la descripción y análisis estadístico de una población, que resume y presenta datos obtenidos de la población o de una muestra, mediante métodos adecuados”*. (Nolberto & Ponce, 2008, p. 17).

En forma general, el rol de la Estadística Descriptiva es resumir lo medular de la información recopilada (Ferreiro, 1998), ya que consiste, sobre todo, en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describir los mismos sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos, como tales:

- Describir información. Esto es válido para el conjunto de datos descrito sólo si y se realiza mediante:
 - i) Tablas de frecuencias y/o porcentajes. ii) Gráficos
 - iii) Medidas que resumen la información, como media o promedio, moda, mediana, desviación estándar, coeficiente de variación, etc. De esta manera una gran cantidad de datos pueden ser mostrados en forma "resumida" y susceptibles de ser interpretados (Rustom, 2012, p. 11).

1.2.2. Estadística Inferencial

La Estadística Inferencial: está referida a métodos empleados, para inferir algo acerca de una población basándose en los datos obtenidos a partir de una muestra e *“incluye un conjunto de técnicas más complejas, cuyo propósito consiste en: poder efectuar deducciones y proponer evidencias científicas sobre una población origen, a partir de los datos registrados en una muestra concreta”* (Granero, 2016, p. 4).

Por su parte Nolberto & Ponce (2008), la definen como: *“rama de la estadística que estudia el comportamiento y propiedades de las muestras y la posibilidad, y límites, de la generalización de los resultados obtenidos a partir de aquellas a las poblaciones que representan”*. (p. 17).

De acuerdo con Ferreiro (1998), la Estadística Inferencial, extiende las conclusiones obtenidas de la muestra a la población de la que ella es parte. Además, postula modelos estadísticos que se ajusten a los datos obtenidos y comprende el:

“conjunto de métodos estadísticos, que permiten deducir (inferir) cómo se distribuye la población bajo estudio a partir de la información que proporciona una muestra representativa, obtenida de dicha población. Para que la Estadística Inferencial proporcione buenos resultados debe:

1. Basarse en una técnica estadístico-matemática, adecuada al problema y suficientemente validada.
2. Utilizar una muestra, que realmente sea representativa de la población y de un tamaño suficiente”. (Nolberto & Ponce, 2008, p. 19).

1.3. Términos estadísticos básicos

1.3.1. Datos y series estadísticos

Los datos estadísticos: son un conjunto de números referidos a una misma característica, y recogidos de tal modo que puedan ser comparados, analizados o interpretados, por lo que son cálculos

aritméticos realizados, sobre los valores obtenidos en una porción de la población, seleccionada según criterios rigurosos.

Para que sean de utilidad, los datos estadísticos, deben tener dos características básicas:

a- Deben ser pertinentes: deben guardar una relación con el tema en cuestión.

b- Deben ser insesgados: no deben tener deformaciones, provenientes de prejuicios o de errores de los instrumentos empleados.

Una serie de datos estadísticos es un conjunto de mediciones efectuadas a un cierto número de Individuos, objetos, etc.

1.3.2. Universo, población y muestra

1.3.2.1. Universo

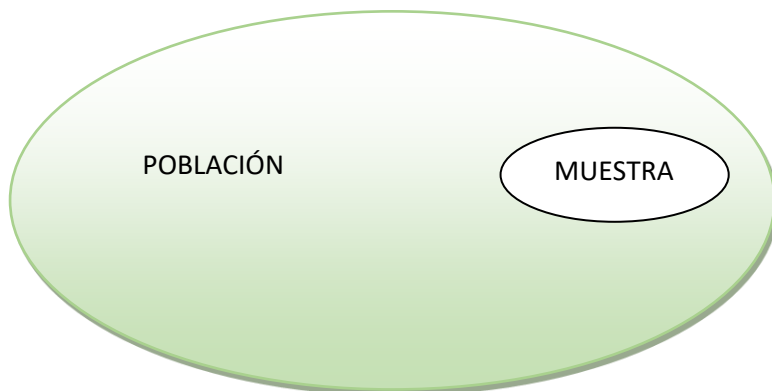
Llamamos universo o colectivo al conjunto de referencia sobre el cual van a recaer las observaciones y representa el conjunto grande de individuos que deseamos estudiar, generalmente, suele ser inaccesible. Es, en definitiva, un colectivo homogéneo que reúne unas características determinadas.

También es entendida: como serie real o hipotética de elementos que comparten características definidas relacionadas con el problema de la investigación.

Por universo se entenderá el conjunto de individuos objeto de interés o estudio. *“La especificación del universo, en general, no es trivial, pues es necesario que no haya ambigüedad, respecto a quien forma parte o no forma parte de este conjunto”*. (Rustom, 2012, p. 12)

1.3.2.2.Población

Metodológicamente, en un trabajo de investigación, se denota a la población como un conjunto definido, limitado y accesible del universo que forma el referente para la elección de la muestra. Es el grupo al que se intenta generalizar los resultados del estudio. Comprende todos los elementos (personas, familias, grupos, objetos, organizaciones, etc.) que presentan características comunes, que se definen a través de criterios establecidos para el estudio. Se debe definir la población con precisión, de modo que sea manifiesto cuándo cierto elemento pertenece o no a esa población. Para el enfoque cuantitativo, la población debe situarse claramente en torno a características de contenido, lugar y tiempo (Gamboa, 2018).



Es todo *“conjunto de elementos, definido por una o más características, de los que gozan todos los elementos que la componen y solo ellos”* (Piña & de Rojas, 2018, p. 197).

Por población se entenderá que *“es el conjunto de datos de una característica medida en cada individuo del universo”*. (Rustom, 2012, p.

11). Así, asociado a un mismo universo se podrán tener varias poblaciones. Para distinguir una población de otra se utiliza lo que se denomina variable. En consecuencia, los diferentes valores que toma una característica se denominan variable.

Otros conceptos teóricos que se deben aclarar en este contexto son el de población finita y población infinita, en tal sentido se define.

- **Poblaciones finitas:** formadas por un número finito de elementos. Se refiere a una población finita cuando la población tiene un tamaño establecido y limitado, esto es, existe un número denotado por N que indica cuántos elementos conforman la población.
- **Poblaciones infinitas:** formadas por un número infinito de elementos. Se refiere a población infinita cuando es teóricamente imposible de contarla, ya que se necesita tiempo y recursos, para observar todos los elementos. Aunque la población sea excesivamente grande, no existe una población infinita de objetos físicos, por lo que entonces se considera infinita, cuando no se puede enumerar en un tiempo razonable.

Arias-Gómez, Villasís-Keever & Miranda-Novales (2016), mencionan otro tipo de población, la población hipotética, que se reconoce cuando el tamaño de la población “*no es posible definirlo en forma precisa, porque se trata de eventos o hechos que aún no han ocurrido*”. (p. 203).

Los parámetros: son las características medibles en la población y se denotan en letra mayúscula

Los parámetros de la población son los siguientes:

- a) Tamaño de la población (N)
- b) Media aritmética (μ)
- c) Varianza (σ^2)
- d) Desviación estándar (σ)

En diferentes contextos de la investigación, al referirse a la población suele denominársela población objetivo, población investigada, población diana, población blanca, población accesible o target group. Estos términos no significan lo mismo, aunque algunos se utilizan como sinónimos, son distintas maneras de definir a la población de acuerdo con el ámbito en el que se quiere estudiar (en ciencias sociales, ciencias biomédicas, ciencias económicas, etc.) (Gamboa, 2018).

1.3.2.3.Muestra

Una muestra es: un subconjunto de una población. Las muestras representativas de una población son útiles, ya que facilitan el manejo de los datos. Se considera representativa de la población, si al escogerla cada elemento tiene la misma probabilidad de salir o de ser escogido.

Entonces, se entiende por muestra a cualquier subconjunto de la población. Existen distintas formas de elegir una muestra (Rustom, 2012).

Las características de la muestra suelen llamarse: **“estadísticos”**. Un estadístico es: un valor numérico, que describe una característica de una muestra. Su valor concreto depende de los valores de la muestra seleccionada en la que es calculado.

Los estadísticos de la muestra son:

- a) Tamaño de la muestra (n)
- b) Media muestral (\bar{X})
- c) Desviación estándar (s)
- d) Varianza muestral (s^2).

1.3.3. Variables

A cualquier característica que puede constatarse en cada individuo de una población, se le denomina: característica aleatoria. Muchas características aleatorias se expresan numéricamente (como: rendimiento académico, número de puntos obtenidos al lanzar un dado y otros), a este tipo de características aleatorias se las denomina variable aleatoria. Una variable aleatoria es: un evento numérico, cuyo valor se determina por medio de un proceso aleatorio.

1.3.3.1. Tipos de Variables

Para representar adecuadamente las poblaciones, es necesario reconocer el tipo de variable que se necesita describir. Podemos distinguir varios tipos de variables:

- **Variables cualitativas:** aquellas que no aparecen en forma numérica, sino como categorías o atributos (sexo, profesión, color de ojos) y sólo pueden ser nominales u ordinales.
- **Variables cuantitativas:** se expresan numéricamente. Se dividen en discretas y continuas.
- **Variables discretas:** una variable es discreta si sus valores se pueden contar; si existe una relación biunívoca con el conjunto

de los números naturales. Existe una unidad mínima que no puede subdividirse. Ejemplo: cantidad de boxeadores, de hijos, cuenta bancaria, etc. No se puede tener media cuenta bancaria o un cuarto de carro.

- **Variables continuas:** los valores de estas variables no se pueden contar, puesto que siempre existe un número entre dos de ellos. Generalmente, se encuentran en los procesos de medición, como peso, altura, temperatura, velocidad.

Autores como: Rustom (2012), las clasifican de 2 formas: cualitativas y cuantitativas. Las cualitativas, se subdividen en: ordinales y nominales y las cuantitativas en: discretas y continuas (Figura 1).

Variables cualitativas

- Ordinales.
- Nominales

Variables cuantitativas

- Discretas
- Continuas

1.3.4. Niveles de medición

La medición es un proceso inherente y consustancial a toda investigación, sea esta cualitativa o cuantitativa. Se miden, principalmente, variables y ello demanda considerar tres elementos básicos: el instrumento de medición, la escala de medición y el sistema de unidades de medición. La validez, consistencia y confiabilidad de los datos medidos dependen, en buena parte, de la escala de medición que se adopte (Coronado, 2007).

Según afirma Bar (2000), la medición se realiza sobre unidades de análisis o sobre las variables. Por lo que hay escalas cualitativas y cuantitativas, existiendo operaciones clasificatorias o sea de ubicación de las unidades de análisis en clases. Así, por ejemplo, para la variable “estado civil” existen las categorías solteros, casados, divorciados, viudos. También están las variables que pueden ser expresarse en intervalos (estatura, peso o nivel de ingresos) o con la intensidad con que la propiedad se manifiesta, propiedad que asume cantidades.

1.3.5. Tipos de escalas

- **Escala nominal:** Las escalas nominales clasifican en función de atributos: (Bar, 2000). Son las más elementales y clasifican a las unidades de estudio en categorías, basándose en una o más características, atributos o propiedades distintivas y observadas, dándole a la categoría un nombre, generalmente, las variables nominales que incluyen dos categorías se denominan dicotómicas (sexo: hombre o mujer; tipo de escuela a la que se asiste: pública o privada) y las variables que tienen tres o más categorías se

denominan: multicotómicas o policotómicas (carrera elegida, raza a la que se pertenece) (Coronado, 2007).

- **Escala ordinal:** Las escalas ordinales sólo poseen orden. Es decir, organizan sus datos a través de las relaciones de igualdad, mayor o menor (Bar, 2000). Las observaciones pueden colocarse en un orden relativo con respecto a la característica que se evalúa, es decir, las categorías de datos están clasificadas u ordenadas, de acuerdo con la característica especial que poseen. Las etiquetas y los símbolos de las características denotan jerarquía (Coronado, 2007).
- **Escala de intervalo o interválicas:** Poseen atributos de orden y distancia o estimación precisa de las unidades (Bar, 2000). Con forme con Coronado (2007), estas escalas reciben también el nombre de cardinales y son más refinadas, ya que pueden establecer orden y jerarquía entre las categorías y las etiquetas o números consecutivos establecen intervalos iguales en la medición. Entonces, se supone que las variables medidas en escalas de intervalo dan una idea de cuánto o de qué tamaño es lo que se está midiendo (Alarcón y Muñoz, 2008).
- **Escalas proporcionales o racionales:** Son las únicas que cuentan con las tres propiedades y se constituyen en verdaderas series numéricas (Bar, 2000). También se llaman escalas de cocientes y tienen las propiedades de las ordinales y de intervalos, pero además, el cero es real y no arbitrario, es absoluto. Es decir, el cero representa la ausencia de la característica en cuestión (Coronado, 2007).

PRÁCTICA

1- Indique en el espacio en blanco cuál distribución es cualitativa y cuál cuantitativa.

- a) Deportistas, por nacionalidad: _____
- b) Entrenadores de béisbol, por sueldo mensual: _____
- c) Jugadores de un equipo, por mes de nacimiento: _____
- d) Jugadores de baloncesto, por estatura: _____
- e) Nadadores, por color de piel: _____

2- Señale si las variables que se presentan son discretas o continuas.

- a) N° de hermanos, por deportistas: _____
- b) Peso en Kg, de los boxeadores _____
- c) Sueldo bruto, de los entrenadores de fútbol: _____
- d) Compras de tipo de camisetas para equipo de béisbol: _____
- e) N° de lesionados, de en un partido de fútbol: _____

3- Señale cuál de las siguientes proposiciones estudia la Estadística Descriptiva y cuál la Inferencial.

- a) Presentar cuadros estadísticos: _____
- b) Prueba en un laboratorio, de la eficiencia de un nuevo tipo de suplemento dietético: _____
- c) Realizar gráficos estadísticos: _____
- d) Estimación de la proyección, de jugadores de tenis para el año 2035: _____
- e) Proyección de la construcción de estadios para los siguientes años: _____

4- Dé ejemplos de 5 variables discretas, 5 variables continuas y 5 variables cualitativas.

5- Se desea promocionar en el mercado un nuevo sabor de bebida refrescante en un estadio de béisbol. Se realiza un test de aceptación

de dicho sabor, en una muestra 20 asistentes al estadio. Se utiliza una escala de 8 puntos, para medir el grado de aceptación, siendo el 8 el máximo valor de aceptación. Las puntuaciones ofrecidas por los atletas son: 6, 8, 7, , 7, 6, 7, 3, 8, 7, 6, 8, 5, 4, 7, 8, 5, 7, 6, 7.

Ahora, responde a las siguientes interrogantes:

a) ¿Cuál es la población? b) ¿Cuál es la muestra? c) ¿Es cualitativa o cuantitativa? d) ¿Cuál es la variable? e) ¿De qué tipo es la variable?

ESTADÍSTICA E INVESTIGACIÓN

2.1. La estadística y la investigación

La metodología y el proceso, que sigue toda investigación facilita la generación de un nuevo conocimiento científico, por la que la investigación y el desarrollo comprenden, generalmente un trabajo creativo llevado a cabo de una forma sistemática, para incrementar el volumen de conocimiento, (el conocimiento del hombre, por ende, de la cultura y la sociedad), así como el utilizar esos conocimientos para crear nuevas formas de aplicarlo (Abello, 2010).

Es dentro de la investigación y el desarrollo, donde la Estadística, relacionada con el método científico y concebida como la ciencia de la recolección y análisis de los datos para la toma decisiones y la encargada de transformar esos datos en información, toma un papel beligerante, apoyando la investigación, en todas las disciplinas del saber humano (Ferreiro, 1998).

Anteriormente, el término estadística, según lo señala Rivera (2017), solo:

[...], era aplicado para asuntos de gobierno, sin embargo, hoy en día se hace extensivo su uso en todas las disciplinas del conocimiento y cómo no en las facetas cotidianas de la vida humana, porque nuestro actuar se basa en tomar decisiones permanentes. Así, utilizamos estadística para la Biología, los

negocios, la Medicina, Sociología, la enseñanza, pero fundamentalmente es una herramienta básica del proceso de investigación a todo nivel (p. 6).

En la actualidad, es ampliamente reconocida la incidencia de la Estadística Aplicada al desarrollo de investigaciones. Es cada vez más utilizada en la recopilación y análisis de datos referidos a conjuntos lo más numerosos posible, donde destacan la variabilidad y la incertidumbre, dándole un papel protagónico, a esta ciencia para convertirse en algo fundamental, para tomar decisiones acertadas (Gamboa, 2017).

Gil (2003), afirma que dentro del proceso investigativo la Estadística se convierte en una herramienta útil que ofrece técnicas y procedimientos en varias etapas de la investigación. Por ejemplo:

- **En planteamiento del problema y formulación de hipótesis:**

Desde esta perspectiva, el investigador, al formular el problema, debe utilizar la Estadística, para poder determinar el tipo de datos que necesita recoger, las técnicas más adecuadas para recolectar esos datos y de cuáles procedimientos estadísticos se valdrá, para el análisis de esos datos. Al igual, que la formulación de la hipótesis no puede hacerse de espaldas a la Estadística, ya que es necesaria para su contrastación.

- **Diseño de investigación:** La estadística forma parte esencial de los diseños de investigación experimentales, ya que se deben contemplar los aspectos que constituyen el experimento y el procedimiento estadístico, que hará posible la interpretación de los resultados. También, la Estadística es fundamental, cuando el diseño incluye la selección de sujetos y en la elaboración de los instrumentos.
- **Análisis de los datos:** Sobre la base de razonar sobre un gran número de observaciones, a la estadística le corresponden las tareas de organización, descripción, análisis y presentación de datos, acerca de las muestras estudiadas y la generalización de los resultados a las poblaciones de donde las muestras se extrajeron. En esta etapa de la investigación y dentro del análisis estadístico de los datos, es de suponer que se debe hacer una descripción de los datos, el descubrimiento de regularidades y la inferencia de características a conjuntos más amplios.
- **Obtención de conclusiones y redacción del informe:** La Estadística proporciona herramientas que formalizan procedimientos para obtener conclusiones que deben incluir resultados estadísticos, mediante la presentación de tablas, cuadros, recogiendo medias, porcentajes, coordenadas, correlaciones o cualquier otro tipo de estadístico.

2.2. La medición en las Ciencias Sociales

Del Canto y Silva (2013), señalan que, a través del tiempo, el ser humano ha estado en la búsqueda constante de nuevos conocimientos, siendo la investigación cuantitativa una de las vías para obtenerlo, teniendo su

soporte en la indagación, a través de elementos cognitivos y en datos numéricos extraídos de la realidad, procesados estadísticamente para probar teorías. Sin embargo, en el campo de las Ciencias Sociales, donde la realidad es cambiante e incierta, producto del contexto complejo en la cual se encuentra inmersa, surge el enfoque mixto (cuali-cuantitativo) como vía para que el investigador pueda acercarse a su objeto de estudio y obtener el conocimiento requerido de la realidad social según sus necesidades y expectativas.

La medición es simplemente: “*el proceso de asignarle valores a ciertos eventos de la realidad*” (Alarcón y Muñoz, 2008, p. 125), y en las Ciencias Sociales, generalmente, se realiza a través de escalas que son: “*técnicas de medida de cantidad de una propiedad*”. (Aignerren, 2010, p. 1). Se usan para medir comportamientos sociales que son difíciles de cuantificar, porque están compuestos por características o dimensiones, no observables o si los tienen, están asociados a componentes afectivos (latentes o no manifestadas) y la dificultad de este proceso radica, al menos en dos aspectos: que el valor represente, el evento que se quiere medir, y que el evento sea expresado en toda su complejidad (Alarcón y Muñoz, 2008).

Por su parte, Coronado (2007), manifiesta que medir es: “*estimar la magnitud de cierta propiedad de uno o más objetos con ayuda de un sistema métrico específico: (instrumento de medición, escala de medición y unidades de medición)*”. (p. 104).

La naturaleza de la medición, en las Ciencias Sociales incluye la evaluación: (proceso mediante el cual las cosas se diferencian) y la escala

de medición: (conjunto de los posibles valores, que una cierta variable puede tomar) (Cazau, 2006).

Para concluir, se quiere señalar que definir el acto de medir trae un sinnúmero de dificultades. Así, para Galtung (citado en Bar, 2000), la medición es un proceso de clasificación de unidades de análisis, según alguna característica elegida. En tanto que en Carmines y Zeller (citado en Bar, 2000), la definen como un proceso de vincular conceptos abstractos, con indicadores empíricos, proceso que supone una previa planificación de operaciones, tanto de clasificación como de cuantificación.

Conforme con Aigner (2010), dentro los niveles de medición, en las ciencias sociales existen técnicas para medir actitudes por escalas y entre las más conocidas están:

- La técnica del escalamiento de Likert.
- El diferencial semántico.
- La escala de Guttman.

2.3. Naturaleza de las técnicas estadísticas utilizables en las Ciencias Sociales




Los griegos hicieron la primera contribución al desarrollo del método para descubrir la verdad o aproximarse a ella, por lo que hay que tener presente que el abordaje investigativo se nutre de las fuentes del conocimiento, a través del razonamiento deductivo o del inductivo y se operacionaliza con el uso del método. El razonamiento deductivo e inductivo, es de gran utilidad para la investigación, así, por ejemplo, la deducción permite establecer un vínculo de unión entre teoría y

observación y permite deducir a partir de la teoría los fenómenos objeto de observación, y la inducción, por su parte, conlleva a acumular conocimientos e informaciones aisladas (Dávila, 2006).

RAZONAMIENTO DEDUCTIVO.

TEORÍA  HIPÓTESIS  OBSERVACIÓN  CONFIRMACIÓN

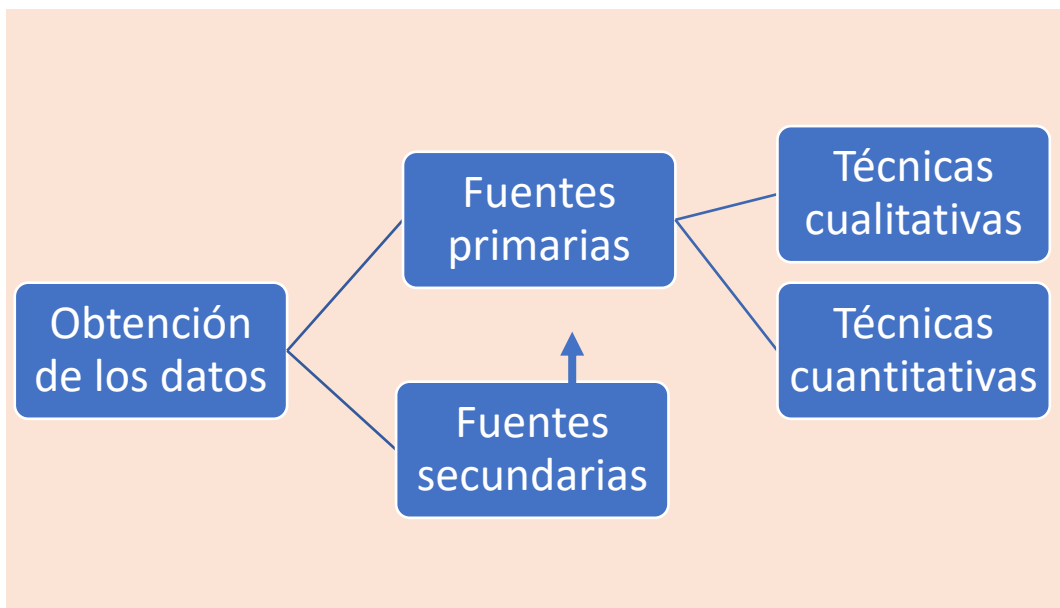
RAZONAMIENTO INDUCTIVO.

OBSERVACIONES  PATRONES  HIPÓTESIS TENTATIVAS 

TEORÍAS

A partir de estos señalamientos, se puede afirmar que toda investigación es selectiva, ya que no es posible capturar totalmente la complejidad de la realidad, por lo que depende de recoger tipos determinados de evidencia a través del prisma de los métodos, cada uno de los cuales tiene sus fuerzas y debilidades (Romo y Castillo, 2002).

La naturaleza de los datos y las técnicas que se utilizan en el proceso de investigación en las ciencias sociales, están íntimamente relacionadas con la forma en que se obtienen los datos y acorde con Lafuente y Marín (2008), éstos se pueden obtener de fuentes primarias y secundarias (Ver Figura 2).



Hasta hace bien poco, según Campo y Gómez (2009), la investigación estaba definida por un marco que la delimitaba en términos exclusivamente cuantitativos. Pero en los últimos tiempos, los paradigmas dominantes en el ámbito de las Ciencias Sociales han entrado en crisis y, una de las consecuencias, ha sido la apertura hacia otras formas de ver y entender la realidad social y, a consecuencia de esto, surgen otras estrategias para estudiarla.

Los autores inmediatamente antes citados, señalan que es conveniente precisar términos como: metodología, método y técnica. Por metodología se entiende:

[...] el conjunto de medios teóricos, conceptuales y técnicos, que una disciplina desarrolla para la obtención de sus fines. Por método, camino que hay que seguir, para acceder al análisis de

los distintos objetos que se pretenden investigar. Las técnicas aluden a procedimientos, de actuación concreta y particular, de recogida de información relacionada con el método de investigación que se está utilizando (p. 1).

Con respecto a las técnicas cualitativas, Carmona (1997), afirma que en las ciencias sociales son utilizadas porque tienen como objeto el estudio del hombre, su conducta y relaciones, aspectos, éstos que no podían ser abordados por el método hipotético-deductivo: (propio de las ciencias de la naturaleza), pero, a la vez, éstas se convierten en técnicas y prácticas de la investigación social. Además, estas técnicas tratan de explicar lo cambiante y lo complejo de la realidad social o de los fenómenos sociales (Mejía, 2003).

Las técnicas cuantitativas de la investigación social: *“brindan un abanico de posibilidades cognoscentes del mundo social, debido a la minuciosidad y el detalle con el que se recogen todas las experiencias vitales, sí como las valoraciones y cosmovisiones del individuo”*. (Mejía, 2003, p. 225).

ESTADÍSTICA DESCRIPTIVA BÁSICA

3.1. Estadística Descriptiva Básica y su aplicación en el ámbito de la Educación Física

La Estadística Descriptiva básica está destinada al estudio de la recolección, análisis e interpretación de datos. Comprende métodos de recolección, descripción, visualización y resumen de los datos, que pueden ser presentados en forma numérica o gráfica.

El uso de la Estadística se ha extendido considerablemente, de modo tal de hacer inteligibles los datos y en consecuencia tomar decisiones en una pluralidad de disciplinas, entre las que se encuentra la Licenciatura en Educación Física.

Al entender la relación que tienen la Estadística Descriptiva básica y la Educación Física, nos podemos percatar que tiene una función formativa, por cuanto, la Estadística Descriptiva es una disciplina que desarrolla una forma especial de pensamiento y cuando se aplica a la Educación Física, tiene un rol instrumental, por cuanto proporciona herramientas de trabajo al futuro profesional de la Educación Física.

Por otra parte, ayuda a sistematizar, resumir e interpretar observaciones para que resulten útiles para obtener la mejor información. Asimismo, es útil para el profesional que investiga, puesto que le permite conocer si los datos apoyan afirmaciones, o bien si el tratamiento dado a los mismos, por ejemplo, en trabajos científicos publicados, es el adecuado. En

síntesis, el conocimiento de las herramientas que proporciona la Estadística Descriptiva es útil para el licenciado que “*observa, lee, evalúa, investiga y valida*”. (Rosi, 2016, p. 2).

Es de suponer que: la Estadística Aplicada a la Educación Física ayuda a sistematizar, resumir e interpretar observaciones, que resulten útiles para obtener la mejor información, por lo que uno de los propósitos de este capítulo es: proporcionar al profesional actual y al futuro de la Educación Física, una serie de técnicas que le faciliten llegar a conclusiones, a partir de bases objetivas y que ayuden a tomar decisiones en las actividades de investigación, que puedan emprender o poner en práctica conocimientos básicos obtenidos en el desarrollo de su carrera.

3.2. Distribuciones de Frecuencias

3.2.1. Concepto

Es un método de organización y resumen de datos y puede darse como una tabla en la cual se agrupan en Intervalos, los valores de la variable y se registra el número de valores observados correspondientes a cada clase.

3.2.2. Tipos de distribuciones de frecuencias

La tabulación es la manera sistemática de realizar el conteo del número de casos u observaciones (datos), obtenidos en una investigación y es donde se establece cuántos pertenecen a una u otra característica. Los datos agrupados pueden resumirse gráficamente o en tablas. El nombre que reciben los datos ordenados en grupos o categorías es el de **Distribuciones de frecuencias.**

Las distribuciones de frecuencias pueden ser: Distribuciones de frecuencias agrupadas y no agrupadas. En las no agrupadas se tienen la Tipo I.

3.2.2.1. Distribución de Frecuencias Tipo I

Concepto

Son estadísticas que constan de pocas observaciones.

Ejemplo:

Ventas diarias de boletos en una semana en un estadio:

Lunes: \$ 275.00
Martes: \$ 300.00
Miércoles: \$ 315.25
Jueves: \$ 258.87
Viernes: \$ 275.38
Sábado: \$ 324.50
Domingo: \$ 321.20

3.2.2.2. Distribución de Frecuencias de Tipo II

Concepto

Son estadísticas que constan de muchas observaciones, pero la variable toma pocos valores.

Ejemplo:

Edad de 30 Jugadores de fútbol.

Edad	Nº de Jugadores
22	8
23	12
24	6
25	4

3.2.2.3. Distribución de Frecuencias de Tipo III

Concepto

Son estadísticas que constan de muchas observaciones y la variable toma también muchos valores y se presentarán en detalle sus componentes y cómo se construye, porque es muy útil en el resumen de datos en la investigación

Ejemplo: Edad de un grupo de 30 personas que practican natación.

Clases	Frecuencias
20 - 29	5
30 - 39	6
40 - 49	11
50 - 59	8

3.3 Tablas de frecuencias y su construcción.

En las distribuciones de frecuencias para datos agrupados podemos debemos identificar las siguientes características que la conforman:

Ejemplo: de una distribución de frecuencias agrupada.

Peso de 40 boxeadores.

Clases	Frecuencias
118 - 127	3
128 - 137	6
138 - 147	14
148 - 157	9
158 - 167	5
168 - 177	3

Los intervalos de clases

Un símbolo que define una clase, tal como 118 -127 de la tabla anterior se conoce como intervalo de clase. El intervalo puede darse de manera abierta y cerrada.

La Marca de Clase

La marca de clase es el punto medio del intervalo de clase. Se obtiene sumando los números extremos izquierdo y derecho que definen el intervalo de clase y luego dividiendo entre dos. La marca de clase se le llama también punto medio de clase. Por ejemplo: del punto medio del primer intervalo de clase sería: $(118+127)/2= 122.5$

Categoría de clase

Con categoría de clase nos referimos a: la posición u orden de los intervalos de clase. Para la tabla de frecuencias, anteriormente expuesta, la primera clase o categoría comprende las pesos de 118 - 127 libras; la segunda de 128 - 137 y así sucesivamente.

Límites de clase

Los límites de clase son: los números en los extremos de los intervalos de clase. Los números ubicados en el extremo izquierdo son los límites inferiores y los del extremo derecho son los límites superiores. En la tabla anterior los límites inferiores serían: 118, 128, 138, 148, 158, 168 y los límites superiores: 137, 147, 157, 167, 177.

Límites reales de clases

Estos límites se calculan restando y sumando 0.5 a los límites inferiores y superiores cuando se presentan intervalos de clases representados por números enteros, si el intervalo está representado con un entero, con un decimal se resta y se suma de los límites superior e inferior 0.05.

Pasos para construir distribuciones de frecuencias para datos agrupados

- 1- Identificar el mayor y el menor de los números entre los datos registrados, para así encontrar el rango o amplitud de la variable, que se define como la diferencia de los valores mínimos y máximos previamente identificados.

2- Hallar el número de clases, el cual se puede encontrar mediante las siguientes fórmulas: $K = \sqrt{n^\circ \text{ de observaciones}}$ o

$$K = 1 + 3.3 \log N$$

3- Encontrar el tamaño del intervalo mediante la siguiente fórmula:

$$TI = \text{Rango} / K .$$

Ejemplo:

1- Las edades de 40 personas que hacen ejercicios aeróbicos en un gimnasio son:

56	78	62	37	54	39	62	60
28	82	38	72	62	44	54	42
42	55	57	65	68	47	42	56
56	56	55	66	42	52	48	48
47	41	50	52	47	48	53	68

a) Agrupe estas edades en una distribución de frecuencias.

Solución:

1 - Determinación del rango = $82 - 28 = 54$

2 - Cálculo del N° de clases = $K = 1 + 3.3 * \log(40) = 6.28 \cong 6$

3 - Identificación del tamaño del intervalo = $\text{Rango} / N^\circ \text{ de clases} = 54/6 = 9$

Luego se tiene:

a) Distribución de frecuencias absolutas

Intervalo	Conteo de casos	Frecuencias absolutas	Frecuencias relativa (*)
28 - 36	I	1	0.025
37 - 45	IIII IIII	9	0.225
46 - 54	IIII IIIII II	12	0.300
55 - 63	IIII IIIII I	11	0.275
64 - 72	IIII	5	0.125
73 - 81	I	1	0.025
82 - 90	I	1	0.025

(*) La frecuencia relativa de clase, se obtiene al dividir la frecuencia absoluta de clase entre el total de datos.

3.4 Representación gráfica de distribuciones de frecuencias

3.4.1 Histograma

El desarrollo del histograma, se atribuye al estadístico francés A. M. Guerra.

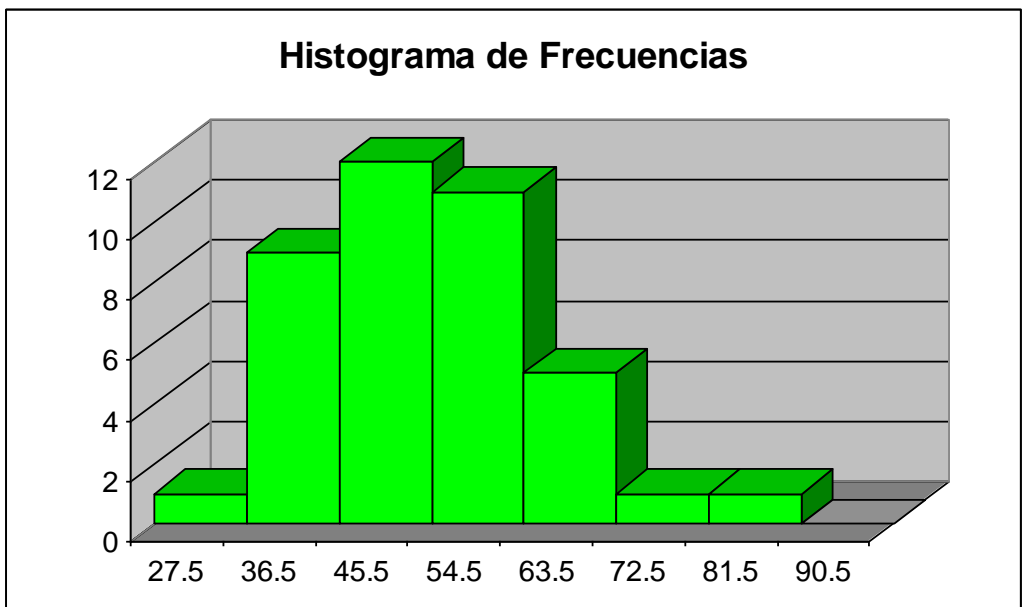
Cuando se estudian variables continuas; las frecuencias se representan mediante áreas o superficies, pues la escala en la que están medidas las mismas permiten tomar valores en cualquier punto del eje horizontal del gráfico.

El histograma de frecuencias es una representación simplificada de una gran masa de datos y debe ser parte del inicio de cualquier análisis más sofisticado. También permite la identificación de valores típicos y atípicos de una distribución.

Un histograma, es un resumen gráfico de la variación de un conjunto de datos. La naturaleza gráfica del histograma, nos permite ver pautas que son difíciles de observar en una simple tabla numérica.

El histograma es un gráfico de barras continuas para ello es necesario restar a cada uno de los límites inferiores -0.5 y sumar a los límites superiores $+0.5$ lo que nos representaría la base de cada barra y la altura de la barra es la frecuencia absoluta de cada intervalo.

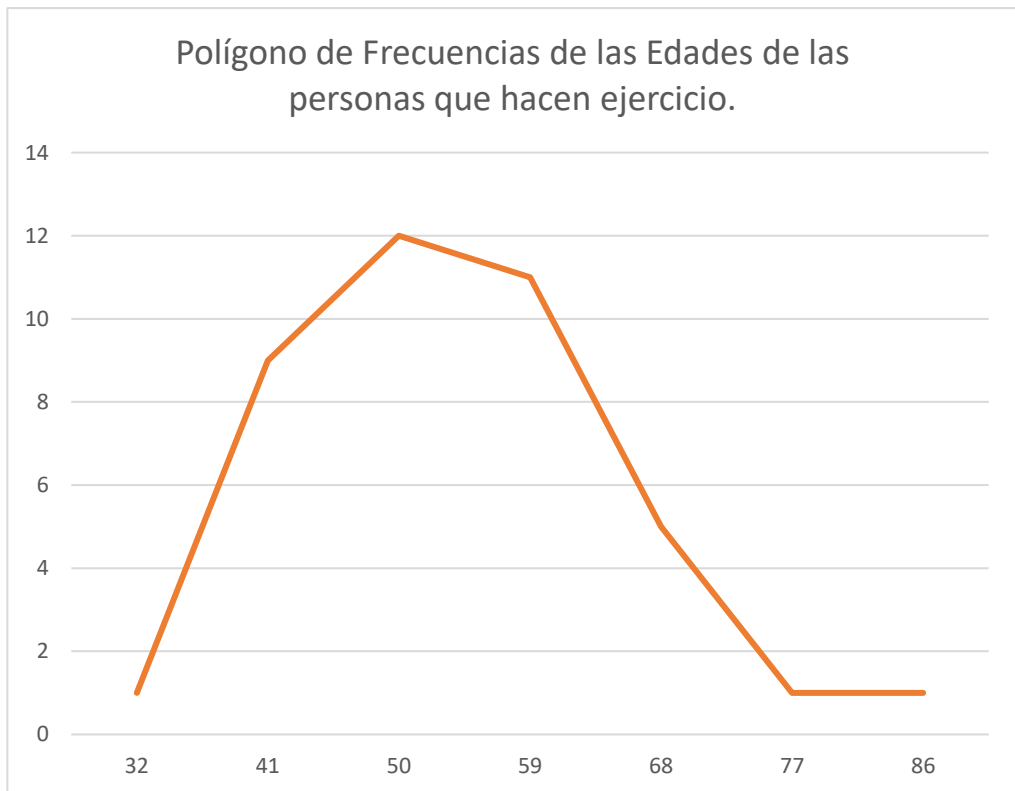
Seria para los límites inferiores $28-0.5=27.5$; $37-0.5 =36.5$ y para los superiores $36+0.5=36.5$; $45+0.5 =45.5$ y así para todos los límites superiores



3.4.2. El polígono de frecuencias

Se utiliza para representar distribuciones de frecuencias (relativas o absolutas). Consiste en representar la gráfica de una función que una por segmentos las alturas correspondientes a los puntos medios de cada intervalo.

Para graficar un polígono debo de encontrar los puntos medios de cada intervalo de clase el cual se encuentra sumando los límites inferiores y superiores de cada clase y dividiendo entre dos Ejemplo: $(28 + 36)/2 = 32$; $(37 + 45)/2 = 41$ y así con cada intervalo de clase.



Problemas Propuestos.

1. Estas son Las edades de 100 personas que se presentaron a un concurso de aeróbicos:

38	51	32	65	25	28	34	12	29	43
71	62	50	37	18	24	19	47	81	53
16	62	50	37	14	17	75	94	16	25
55	38	46	16	72	64	61	33	59	21
13	92	37	43	58	52	88	27	74	66
63	28	36	19	56	84	38	6	42	50
98	51	62	13	17	43	47	54	58	26
12	42	34	68	77	45	60	31	72	23
18	22	70	34	15	59	20	68	55	49
33	52	14	40	38	54	50	11	41	76

Agrupar los datos en clases y presentar dichos datos en una tabla de intervalos de clase.

2. En una cierta ciudad, se registra el número de nacimientos ocurridos por semana durante las 52 semanas del año, siendo los siguientes los datos obtenidos:

6	4	2	8	18	16	10	6	7	5	12	8	9
12	17	11	9	16	19	18	18	16	14	12	7	10
3	11	7	12	5	9	11	15	9	4	1	6	11
7	8	10	15	3	2	13	9	11	17	13	12	8

a) Construya la distribución de frecuencias, grafique el polígono y el histograma.

3. Las edades de veinte jóvenes son: 12, 13, 14, 10, 11, 12, 11, 13, 14, 12, 10, 12, 11, 13, 12, 11, 13, 12, 10 y 15. Organiza los datos en una tabla de frecuencias.

¿Qué porcentaje de chicos tienen 12 años?

¿Cuántos chicos tienen menos de 14 años?

4. Las estaturas en cm de 30 jugadores de fútbol son:

170	176	180	185	170	176
162	162	185	170	176	180
160	167	167	180	162	176
185	176	167	170	185	176
187	170	162	176	170	167

Construya la distribución de frecuencias, grafique el polígono y el histograma.

5- En un club deportivo, donde se practican diferentes disciplinas se aplica un test de aptitud para seleccionar a sus miembros. La cantidad de preguntas contestadas, en forma correcta en una prueba de aptitud fueron las siguientes:

112	73	126	82	92	115	95	84	68	100
72	92	128	104	108	76	140	119	98	85
69	76	118	132	96	91	81	113	115	94
97	86	127	134	100	102	80	98	106	106
107	73	124	83	92	81	106	75	95	119

a) Agrupe estos datos en una distribución de frecuencias y grafique el polígono.

3.5 Los cuadros y gráficos estadísticos.

Los cuadros

Un cuadro estadístico es una lista lógica de datos cuantitativos interrelacionados, o sea es el arreglo ordenado de los datos para facilitar la lectura e interpretación de estos.

Representan la síntesis de los pasos de recopilación, elaboración y análisis de los datos. Se estructuran desde el punto de vista de la utilidad que preste al usuario común.

Las formas básicas de presentación de información son cuatro: textual, semitabular, tabular y gráfica. De estas formas, la más usada es la tabular, siguiéndole en orden de importancia la gráfica.

Los componentes del cuadro estadístico son nueve, sección y número del cuadro, título, encabezamiento, columna matriz, cuerpo del cuadro, columnas, nota, llamadas, fuente de información.

a) Sección y número del cuadro: No es más que el número, código o elemento que ubica el cuadro en la publicación.

b) Título: descripción que precede al cuadro, debe redactarse en forma breve y se recomienda utilizar la forma de pirámide invertida.

Debe responder a las interrogantes: ¿qué, dónde, cómo y cuándo?

¿**Qué?**: trata al hecho observado.

¿**Dónde?**: identifica al lugar que corresponde la información.

¿**Cómo?**: trata el orden que se presenta la información, la cual debe comenzar por el encabezamiento identificado por la

preposición: “por” y debe seguir con la información de la columna matriz identificada con la preposición: “según”.

¿Cuándo?: se refiere al tiempo en que se cubre la información.

c) Encabezamiento: recibe este nombre la serie de casillas situadas en la parte superior del cuadro y en la cual se indican una o varias clasificaciones de los datos.

d) Columna Matriz: colocada en el lado izquierdo es de hecho la primera columna del cuadro. En ella se indican tanto las características principales como las dependientes.

e) Cuerpo del cuadro: es el conjunto de columnas y líneas que contiene el cuadro en orden vertical y horizontal, en él se colocan los datos sobre el hecho observado y las cifras deben ser alinearse a la derecha, unidades debajo de unidades, decenas debajo de decenas, etc. Los totales se colocan en la parte superior, horizontalmente y a la izquierda.

f) Columnas: están localizadas en el cuerpo del cuadro y nunca deben aparecer en blanco.

g) Nota: está destinada a ofrecer conceptos o definiciones que aclaren el contenido de los cuadros o a indicar la metodología utilizada en la investigación.

h) Llamada: contiene información de carácter específico que se aplica a determinada parte del cuadro. Se indica en el cuadro con el símbolo (), Cuando va al lado de un número se identifica con una letra y debe ir antes de ésta. Si va al lado de una palabra se pone en número y al final de ella.

i) Fuente de Información: Indica la publicación o entidad responsable de la información.

Estructura del Cuadro Estadístico:

	1-N° del Cuadro	2-Titulo	3- Encabezamiento	
4- Columna Matríz	5 Cuerpo del	cuadro		
		6- columnas		

7-Nota

8-Llamada

9-Fuente de Información

Ejemplo:

Señale los errores que contiene el cuadro y reestructúrelo siguiendo los pasos establecidos para la construcción de cuadros estadísticos.

Según sexo de población, 2000 deportistas universitarios en Panamá 2018.

Lugar de Procedencia	Total	Masculino	Femenino
Total...	2000	836	1164
Urbana	240		140
Rural	360	160	
Suburbana		576	824

Fuente: Instituto para la Formación de Recursos Humanos.

No tiene número que lo identifique.

- Título, mal redactado
- El título no está en forma de pirámide invertida
- En la columna matriz, la palabra total está mal ubicada
- Falta el rayado
- Faltan cifras
- Está cerrado en el lado lateral
- Fuente incorrecta

Corrección

¿Qué?: POBLACIÓN DE DEPORTISTAS UNIVERSITARIOS.

¿Dónde?: EN PANAMÁ.

¿Cómo?: POR SEXO, SEGÚN LUGAR DE PROCEDENCIA.

¿Cuándo?: AÑO 2018.

CUADRO N°1. POBLACIÓN DE DEPORTISTAS UNIVERSITARIOS EN PANAMÁ POR SEXO, SEGÚN LUGAR DE PROCEDENCIA. AÑO 2018.

Lugar de residencia	Total	Sexo	
		Masculino	Femenino
Total...	2000	836	1124
Urbana	240	100	140
Rural	360	160	100
Suburbana	1400	576	824

Fuente: Instituto Panameño de Deportes.

Para las Normas APA

Tabla 1:

Población de Deportistas Universitarios en Panamá por Sexo según Lugar de Procedencia, Año 2018.

Lugar de residencia	Total	Sexo	
		Masculino	Femenino
Urbana	240	100	140
Rural	360	160	100
Suburbana	1400	576	824
Total...	2000	836	1124

Fuente: Instituto Panameño de Deportes

3.6 Los gráficos

Los gráficos constituyen uno de los medios más conocidos para la presentación y el análisis de la información Estadística, ya que permiten una visión más clara, rápida y atractiva de lo que se está presentando, aunque no nos den una expresión exacta de las cifras como los cuadros estadísticos.

La presentación gráfica de información numérica se basa en: un sistema de coordenadas, sobre la cual se ubican los datos.

El gráfico debe constar de un sistema de coordenadas, escala, título de la escala, título del gráfico, fuente y leyenda. Hay que señalar que, no existen reglas estrictas en la elaboración de gráficos.

Gráficos circulares

Gráfico de pastel

Es un círculo dividido en sectores, en el cual la medida angular de cada sector es proporcional a la magnitud del valor que representa y el círculo completo equivale a la suma de todos los componentes representados.

Se utiliza para mostrar las proporciones entre las variables que componen o integran un todo. Se puede construir con cifras absolutas o cifras relativas, ya que el círculo completo representa el 100%.

Se recomienda que las cifras absolutas de la serie a representar, se ordenen según su magnitud y luego se reduzcan a valores relativos o porcentuales y que en el gráfico los sectores también deben aparecer según el ordenamiento anterior.

Una vez trazada la circunferencia, se debe trazar un radio verticalmente al centro del círculo, de manera que sirva de partida y desde él con un transportador se midan los porcentajes de cada uno de los sectores o componentes del gráfico.

Dentro de cada sector se anota su valor porcentual y de existir espacio también se rotula dentro el significado del sector, de lo contrario puede ubicarse fuera pero próximo al segmento a identificar.

Es importante resaltar que el círculo representa 360° y que es necesario convertir la distribución porcentual a grados y que esto se obtiene por regla de tres simple.

Ejemplo:

La Población de 2000 deportistas universitarios se dividen en extracción urbana 240 estudiantes; suburbana 1400 estudiante y rurales 360 estudiantes. Con esta información realice un gráfico circular o de pastel.

Solución:

Tenemos que la población total es de 2000 estudiantes, esto nos representa el 100%. Luego por regla de tres simple calculamos los porcentajes:

% Urbana	Sub-urbana	Rural
100% → 2000	100% → 2000	100% → 2000
X 240	X 1400	X 360
$X = \frac{240 \times 100}{2000}$	$X = \frac{1400 \times 100}{2000}$	$X = \frac{360 \times 100}{2000}$
X = 12%	X = 70%	X = 18%

Transformación a grados

Urbana

$360^\circ \rightarrow 100\%$

X 12%

$$X = \frac{12 \times 360}{100}$$

X = 43.°

Sub-urbana

$360^\circ \rightarrow 100\%$

X 70%

$$X = \frac{70 \times 360}{100}$$

X = 252°

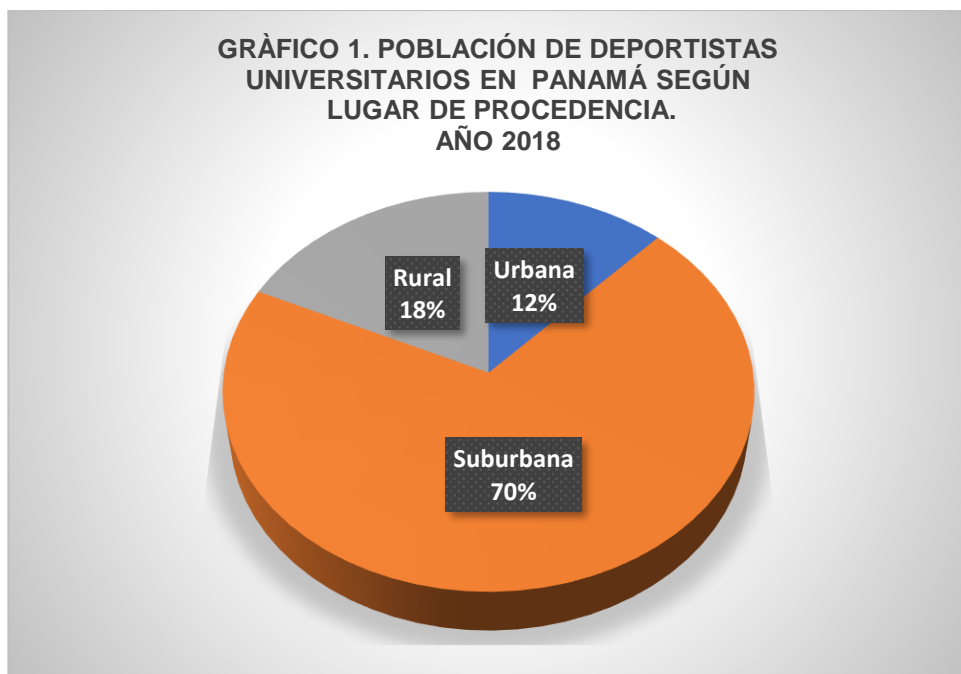
Rural

$360^\circ \rightarrow 100\%$

X 18%

$$X = \frac{18 \times 360}{100}$$

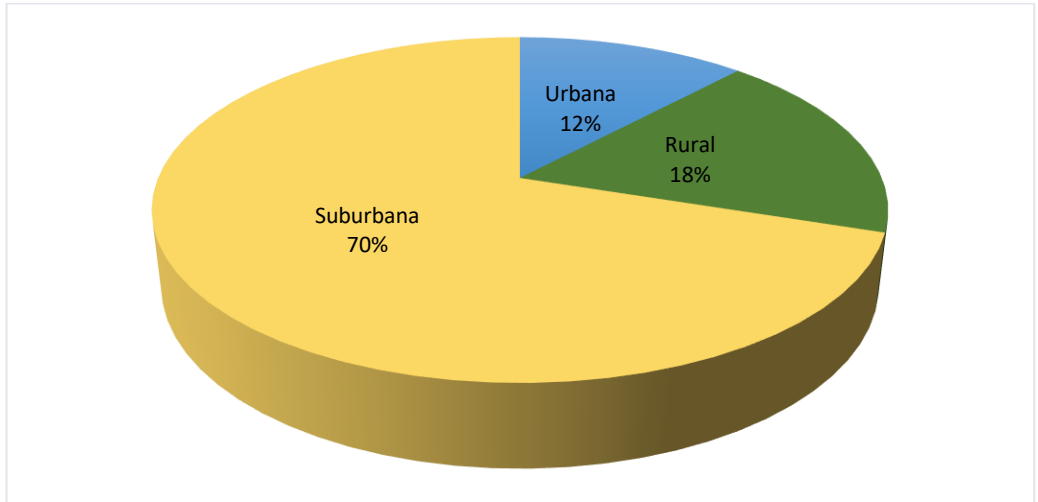
X = 65°



Fuente: Instituto Panameño de Deportes.

Para el Modelo de Normas APA

Figura 1. Población de Deportistas Universitarios en Panamá por Lugar de Procedencia Año 2018.

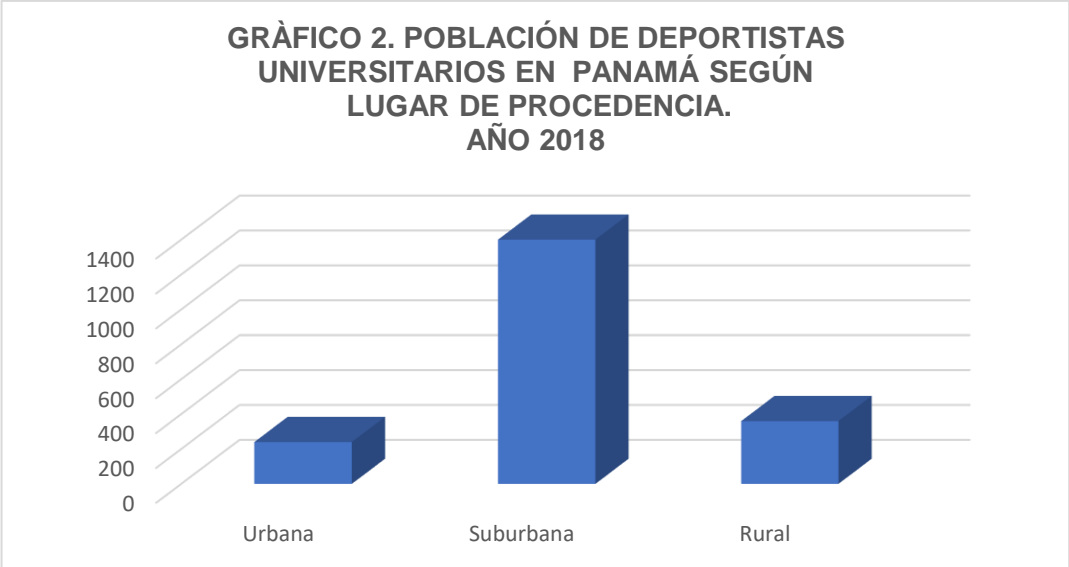


Fuente: Instituto Panameño de Deportes

Gráficos de barras

Están compuestas por rectángulos de igual anchura y de longitudes variables, que son determinadas por la magnitud de los datos. Es uno de los tipos más simples y quizás los más usados con mayor frecuencia, debido a que su presentación permite una fácil interpretación, pues solo es necesario establecer las diferencias entre las longitudes de los rectángulos.

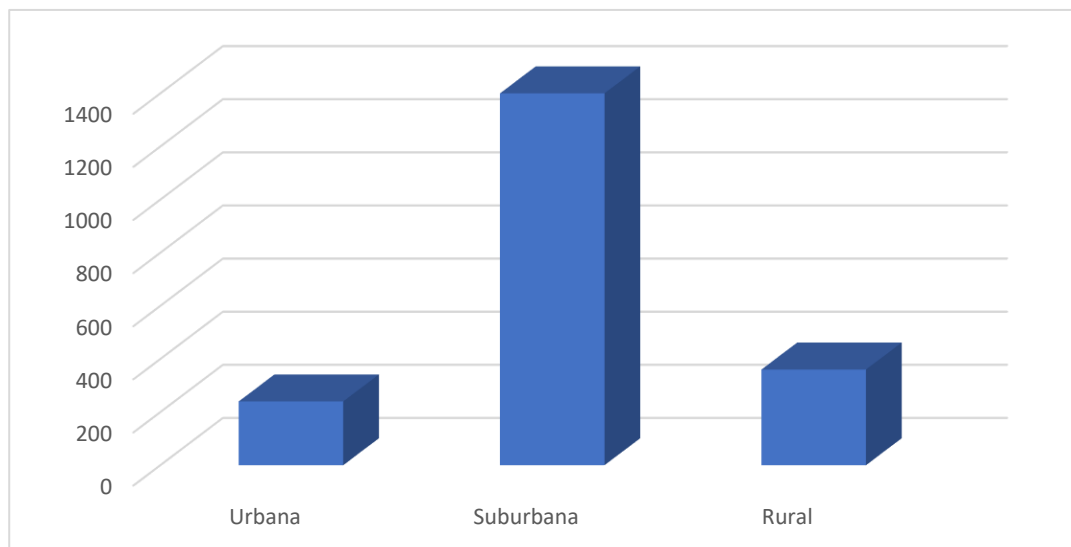
Norma de la Contraloría:



Fuente: Instituto Panameño de Deportes

Para Normas APA

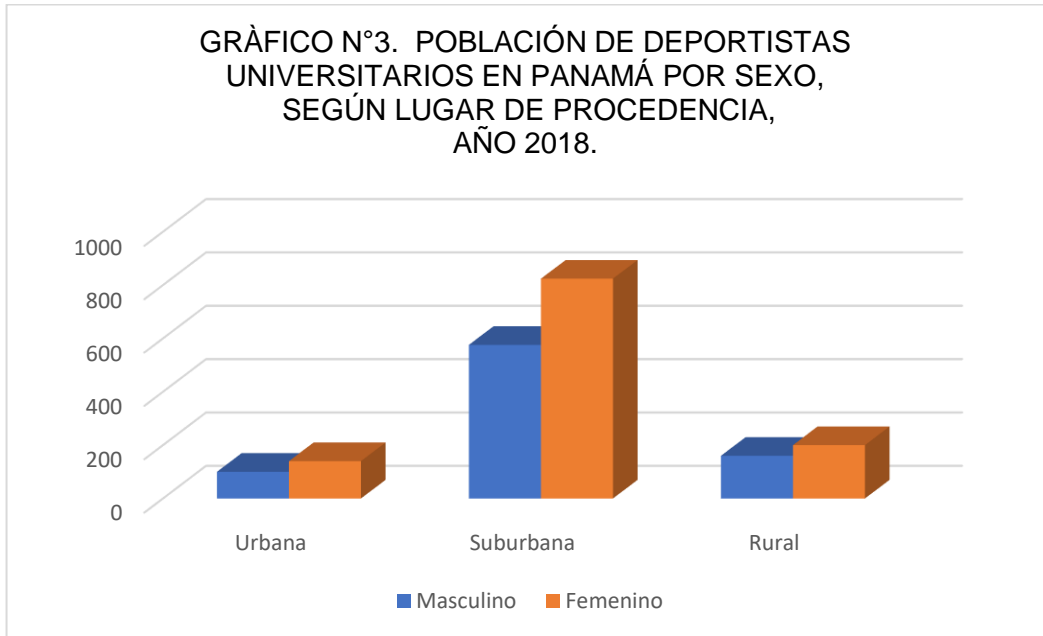
Figura 2. Población de Deportistas Universitarios en Panamá por Lugar de Procedencia Año 2018.



Fuente: Instituto Panameño de Deportes

Gráfico de Barras Comparadas

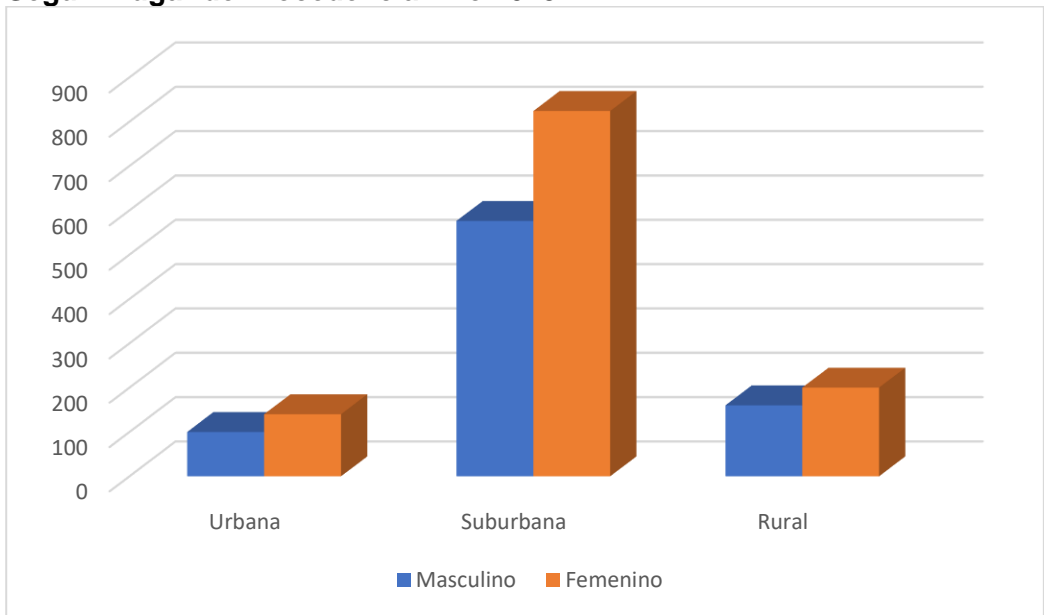
Norma de la Contraloría



Fuente: Instituto Panameño de Deportes

Para Normas APA

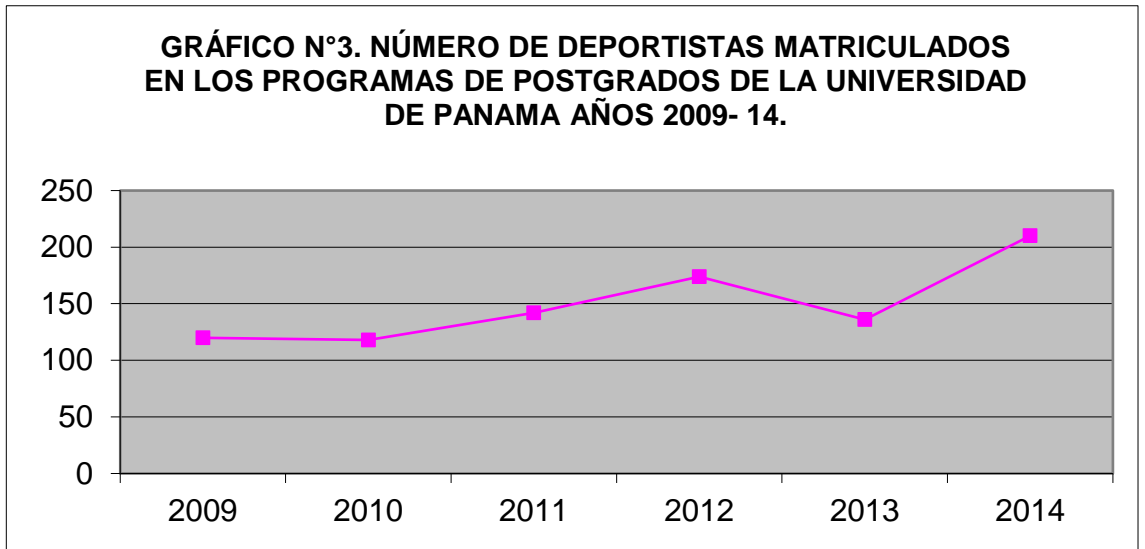
Figura 3. Población de Deportistas Universitarios en Panamá por Sexo Según Lugar de Procedencia Año 2018.



Fuente: Instituto Panameño de Deportes

Gráficos de líneas:

Corresponden a la expresión de una variable en el tiempo.



Fuente: Universidad de Panamá

Práctica de cuadros tablas gráficos y figuras

1- Los siguientes son los números de goles a favor (GF) y goles en contra (GC) de seis equipos de la liga española. Equipos: Barcelona 39 gf y 17 gc; Atlético 36 gf y 8 gc; Real Madrid 35 gf y 18 gc; Real Sociedad 32 gff y 19 gc; Levante 29 gf y 29gc y Villarreal 29 gf y 20 gc. Construya el cuadro estadístico., la tabla el gráfico y la figura que más se ajusta a los datos.

2- Escriba los errores que presenta el cuadro estadístico, reestructúrelo y realice el gráfico que más se ajusta a los datos:

Clasificación según preferencia deportiva en Universidad Estatal en Panamá de 325 estudiantes de acuerdo al sexo

Deporte Preferido		Mujeres	Hombres
Futbol	108	43	65
Voleibol		32	55
Tenis		16	
Gimnasia		25	16
Total...	325		200

3- Escriba los errores que se presentan en el cuadro estadístico, reestructúrelo y construya el gráfico que más se ajusta a los datos.

Ocupación de Jugadores de la Liga Panameña de Fútbol por Región Año 2021

Ocupación	Total	Región de Panamá	Región del Interior
	200		
Sin Trabajo	98	55	
Empresa Privada		38	47
Empleado del Gobierno	17		10

Fuente: Pandeportes

4- Tabla 1.

Preferencia en Deportes en Tres Colegios de la Provincia de Coclé. Año 2020

Deporte	Colegios		
	COLEGIO 1	COLEGIO 2	COLEGIO 3
Futbol	160	155	163
Natación	35	15	35
Voleibol	95	82	79
Gimnasia	8	25	44

Fuente: Ferreterías de Penonomé.

A) Realice los ajustes necesarios a a tabla y grafique la figura que corresponde.

3.7. Las medidas de tendencia central

Tienen como propósito encontrar un valor o medida con la cual podamos comparar datos con otros datos afines, ya que sustituyen al conjunto de observaciones debido a que tratan de resumir en un solo número, la posición o localización de la distribución de los datos.

Las medidas de tendencia central también suelen ser llamadas promedios. Las medidas de posición pueden ser de posición central o no. Las medidas de posición o de tendencia central que caracterizan la distribución son: la media aritmética, la mediana y la moda.

3.7.1. Medidas de tendencia central para datos simples

La media

Es la medida de tendencia central más conocida y comúnmente utilizada, generalmente, es representada por \bar{X} y puede obtenerse sumando un conjunto de puntajes y dividiendo entre estos.

Su fórmula es $\bar{X} = \frac{\sum X_i}{n}$, donde \bar{X} = La media, \sum la suma y n el número total.

Ejemplo:

Encontrar la media de los números: 5 , 8, 3, 2, 10, 2 es

$$\bar{X} = \frac{5+8+3+2+10+2}{6} = \frac{30}{6} = 5$$

La mediana

Se caracteriza en que divide un conjunto de datos ordenados en dos grupos iguales; la mitad de los números tendrá valores menores o iguales

que la mediana, y la otra mitad alcanzará valores mayores o iguales que esta. Para encontrar la mediana es necesario ordenar los valores de menor a mayor o viceversa.

En el cálculo de la mediana se debe atender a las siguientes reglas:

1- Cuando la cantidad de datos es impar o sea n es impar, la mediana es el valor del elemento que está a la mitad.

Ejemplo:

Se tiene la siguiente serie de números ordenados: 1, 2, 5, 7, 8, 11, 14 como se observa el 7 es el elemento que se encuentra en la mitad por tal razón 7 es la mediana de los datos.

2- Cuando n es par o sea la cantidad de datos es par, la mediana es la media de los dos valores que están en el centro. O sea, se suman estos dos valores y los divide entre dos.

Ejemplo:

Diez personas enviadas para entrevistar a 50 estudiantes en cada una de 10 universidades, descubrieron que 18, 14, 15, 12, 8, 21, 7, 11, 16, y 3 de los estudiantes gustaban trotaban con regularidad.

Solución:

Primero tenemos que ordenar los datos y quedarían así: 3, 7, 8, 11, 12, 14, 15, 16, 18, 21 como observamos los dos elementos que están en el centro son 12 y 14 por lo que la mediana sería:

$$\text{mediana} = (12 + 14)/2 = 13.$$

La moda

La moda de una serie de números es aquel o aquellos valores que más se repiten o sea el valor más común. Puede no existir en algunos casos y también puede existir más de una moda.

Ejemplo 1:

El sistema de números: 2, 2, 5, 7, 9, 9, 9, 10, 10, 11 y 12 tiene moda 9.

Ejemplo 2:

La serie de números; 3, 5, 7, 10, 12, 13, 16 y 17 no tiene moda, ya que no se repite ningún valor.

Ejemplo 3:

La serie de números: 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 8 y 9 tiene dos modas 4 y 7.

3.7.2. Medidas de tendencia central para datos agrupados

La media

$$\bar{X} = \frac{\sum f \text{ pm}}{n}, \text{ donde } f = \text{frecuencia; pm} = \text{punto medio y } n = \text{número de datos}$$

Procedimiento de cálculo

Lo primero que tenemos que hacer es calcular los puntos medios de cada clase y luego multiplicarlos por la frecuencia correspondiente a cada clase, por último, realizar la suma de todas esas multiplicaciones y dividir entre el total de datos.

Intervalo	f	pm	f pm
38 – 42	9	40	360
43 – 47	4	45	180
48 - 52	7	50	350
53 – 57	10	55	550
58 - 62	6	60	360
63 – 67	3	65	195
68 – 72	1	70	70
			2065

Luego se tiene:

$$\bar{X} = \frac{\sum f \text{ pm}}{n} = \frac{2065}{40} = 51.62$$

La mediana para datos agrupados

El cálculo de la mediana para datos agrupados se efectúa mediante la siguiente fórmula:

$$Md = L_1 + \frac{(N/2 - \sum f_a)}{f_{\text{mediana}}} C \quad \text{donde,}$$

L_1 = Límite real Inferior del intervalo que contiene la mediana

$N/2$ = N° total de datos dividido entre dos

$\sum f_a$ = Sumatoria de las frecuencias por debajo de la clase que contiene la mediana.

f_{mediana} = Frecuencia del intervalo que contiene la clase mediana.

C = Amplitud del intervalo.

Para calcular la mediana para datos agrupados lo primero que tenemos que hacer es ubicar el intervalo donde esta cae. Para este fin dividimos la cantidad de datos entre dos o sea $n/2$, lo cual nos dará la posición o el intervalo en que esta se encuentra.

Para la distribución de frecuencias que hemos trabajado tenemos: $N/2 = 40/2 = 20$,

$$\begin{aligned}Md &= L_1 + \frac{(N/2 - \sum f_a)}{f_{\text{mediana}}} C \\Md &= 47.5 + \frac{(20 - 13)}{7} 5 \\&= 47.5 + 5 \\&= 52.5\end{aligned}$$

Cálculo de la moda para datos agrupados:

La fórmula es la siguiente:

$$\text{Moda} = Li + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c \quad \text{donde:}$$

Li= Límite real inferior de la clase que contiene la moda

Δ_1 = Exceso de la frecuencia modal sobre la clase contigua inferior

Δ_2 = Exceso de la frecuencia modal sobre la clase contigua superior

C = Tamaño del intervalo.

En la distribución de frecuencia que tenemos, lo primero a ubicar es cuál es el intervalo que presenta la mayor frecuencia y tenemos que es el intervalo que va de 53- 58, cuya frecuencia es 10 y es la más alta.

Luego calculamos los valores de Δ_1 y Δ_2 :

$$\Delta_1 = 10 - 7 = 3 \quad \text{y} \quad \Delta_2 = 10 - 6 = 4$$

Por lo que tenemos:

$$\begin{aligned} \text{Moda} &= Li + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) c \\ &= 52.5 + \left(\frac{3}{3 + 4} \right) 5 \\ &= 52.5 + 2.14 \\ &= 54.64 \end{aligned}$$

Cuándo usar cada medida de tendencia central

La media

Es utilizada como o con datos de intervalo o de razón, además puede calcularse sin ordenar los datos y todo conjunto de datos posee sólo una media. Sirve para realizar procedimientos estadísticos como comparación de las medias a partir de varios conjuntos de datos. No es útil cuando poseemos intervalos de clases abiertas.

La mediana

Puede ser fácilmente calculada una vez que los datos se hayan ordenado en categoría numérica. Es adecuado para usarse con datos ordenados; cuando la distribución contiene clases abiertas y cuando existen valores excepcionalmente grandes o pequeños.

La moda

Puede emplearse cuando la distribución presenta clases abiertas. Su importancia radica en que puede utilizarse para datos cualitativos y también cuantitativos. Se usa menos que la mediana y la media y presenta el inconveniente de que algunas veces no existe el valor modal y otras veces todos los valores son la moda, ya que ocurren el mismo número de veces.

3.8 Medidas de dispersión

Juegan un papel muy importante en la Estadística, pues si los fenómenos no se repitiesen la Estadística no tendría razón de ser. El hecho de que los fenómenos se repitan hace que los mismos muestren variaciones de

mayor o menor intensidad. Así es tan importante conocer un promedio como conocer la variabilidad de los que están a su alrededor.

La importancia de las medidas de variabilidad se hace más clara, si se nota que en la práctica sucede que varios conjuntos de datos presentan una misma media y su dispersión es muy diferente.

Ejemplo:

En una escuela se calculan los promedios finales de los estudiantes de noveno grado para adjudicar los puestos de honor. Cabe señalar que solo existe una sola beca y debe otorgársele al primer puesto de honor. Ayude al Director del Centro a tomar su decisión calculando el promedio para encontrar quién ocupa el primer puesto de honor.

ESTUDIANTE 1		ESTUDIANTE 2	
Español	4.5	Español	4.8
Matemáticas	4.3	Matemáticas	4.0
Ciencias Nat.	4.5	Ciencias Nat.	4.0
Ciencias Soc.	4.4	Ciencias Soc.	4.6
Inglés	4.5	Inglés	4.2
Ed. Física	4.3	Ed. Física	5.0
Agricultura	4.6	Agricultura	4.5
$\bar{X} = 31.1/7 = 4.4428$		$\bar{X} = 31.1/7 = 4.4428$	

Se observa que ambos estudiantes poseen el mismo promedio se procede entonces a calcular la varianza y desviación estándar

Luego calculamos la varianza y desviación estándar.

$\sum (\bar{X} - X)^2$	$\sum (\bar{X} - X)^2$
$(4.5 - 4.4)^2 = 0.01$	$(4.8 - 4.4)^2 = 0.16$
$(4.3 - 4.4)^2 = 0.01$	$(4.0 - 4.4)^2 = 0.16$
$(4.5 - 4.4)^2 = 0.01$	$(4.0 - 4.4)^2 = 0.16$
$(4.4 - 4.4)^2 = 0$	$(4.6 - 4.4)^2 = 0.04$
$(4.5 - 4.4)^2 = 0.01$	$(4.2 - 4.4)^2 = 0.04$
$(4.3 - 4.4)^2 = 0.01$	$(5.0 - 4.4)^2 = 0.36$
$(4.6 - 4.4)^2 = 0.04$	$(4.5 - 4.4)^2 = 0.01$
$S^2 = \frac{\sum (\bar{X} - X)^2}{n - 1} = \frac{.09}{6} = 0.015$	$S^2 = \frac{\sum (\bar{X} - X)^2}{n - 1} = \frac{.93}{6} = 0.155$
$S = \sqrt{0.015} = 0.1224$	$S = \sqrt{0.1555} = 0.3937$

Se tiene entonces que el estudiante I posee una desviación estándar menor, aunque en la práctica del promedio ambos mantenían un promedio igual.

La desviación estándar nos indica que el estudiante 1 mantuvo sus calificaciones más próximas al promedio y es el que debe obtener la beca.

Práctica

1- Las siguientes son las puntuaciones de un grupo de estudiantes de la carrera de Educación Física que participaron de dos modalidades de enseñanza: presenciales y a distancia. Calcule la media, mediana, moda, varianza y desviación estándar para cada una de ambas modalidades.

¿Señale en cuál de los dos grupos la media fue más representativa y por qué?

Grupo presencial:	68	72	71	76	76	80	
Grupo experimental	80	78	82	92	92	94	96

2- En dos secciones de un curso de biometría las distribuciones de las notas obtenidas son:

Sección I		Sección II		Determine para cada sección: a) La media, mediana y moda y señale en cuál la media es más representativa.
Clases	f	Clases	f	
30-39	2	40-49	2	
40-49	5	50-59	4	
50-59	9	60-69	20	
60-69	18	70-79	12	
70-79	10	80-89	5	
80-89	4	90-99	3	

3- Se forman tres grupos de 6 alumnos y a cada uno se le aplica un método de enseñanza. Encuentre la media y la desviación estándar de los tres grupos y explique sus resultados. Los datos del examen son:

Presencial	4,8	7,1	5,4	6,8	8,6	6,2
Clases virtuales	4,9	6,1	5,4	3,6	4,2	2,4
Autodidacta	1,5	6,4	3,9	5,3	2,4	3,1

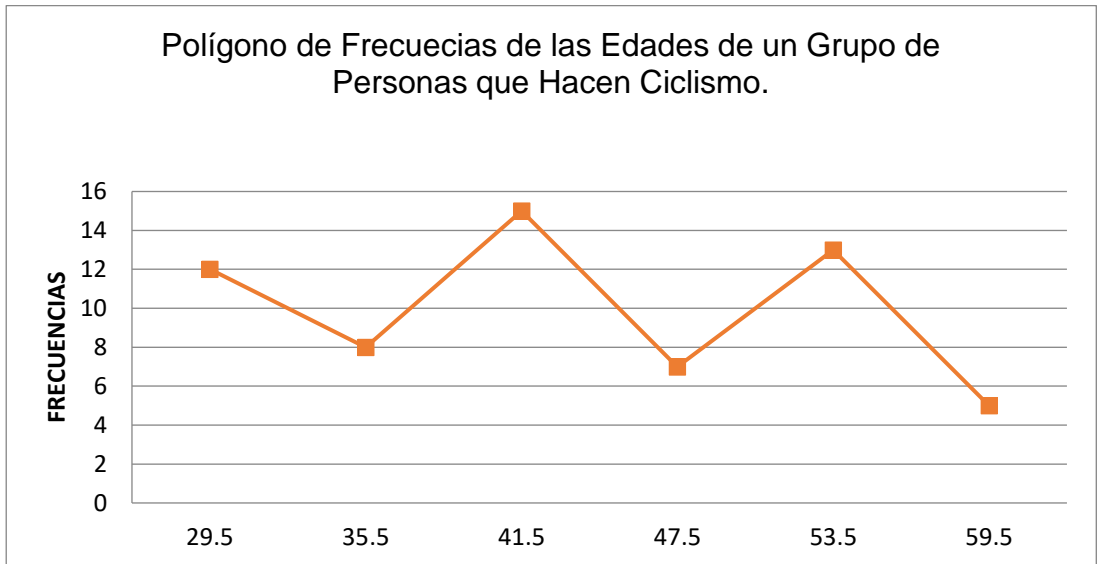
4- El siguiente cuadro muestra la distribución de frecuencia de un grupo de trabajadores no calificados:

Salarios Semanales de 100 Trabajadores no Calificados

Salarios	Frecuencias
140 – 159	7
160 - 179	20
180 – 199	33
200 - 219	25
220 - 239	11
240 - 259	4

- A- Calcular el salario medio de los trabajadores.
- B- ¿Cuál es el salario medio percibido por el mayor número de trabajadores?
- C- ¿Cuál es el salario tope percibido por un empleado que se encuentra el grupo del 50% de los menos favorecidos?
- D- ¿Cuántos trabajadores no llegan a un salario de B/. 180?

5- Dado el siguiente polígono calcule las medidas de tendencia central y de dispersión.



6- Los sueldos mensuales iniciales en dólares de los trabajadores en una Empresa X en el año 2000 fueron: \$ 500, 750, 825, 680, 850, 850 y 780. Para el año 2015 el salario inicial en la misma empresa fue de \$ 620, 675, 900, 750, 780, 825, y 900. Ha mejorado la política salarial de la empresa y se distribuye mejor la escala salarial entre los trabajadores, sustente su respuesta por medio de la medida de tendencia central y de dispersión más indicada

7- Las siguientes distribuciones de frecuencias representa el número de días por año que los jugadores de dos equipos de fútbol han faltado por enfermedad.

<i>EQUIPO UNO</i>		<i>EQUIPO DOS</i>	
<i>Número de días de ausencia</i>	<i>Número de jugadores.</i>	<i>Número de días de ausencia</i>	<i>Número de jugadores</i>
0 - 2	5	0 - 2	8
3 - 5	14	3 - 5	10
6 - 8	12	6 - 8	9
9 - 11	14	9 - 11	14
12 - 14	5	12 - 14	9

- Calcule las medidas de tendencia central y de dispersión para ambos equipos.
- En cuántos días se estima el porcentaje mayor de ausentismo en ambos equipos
- En cuál de los dos equipos el promedio días de ausentismo es mayor

ANÁLISIS DE REGRESIÓN

4.1. Definición

El análisis de regresión y correlación comprende el análisis de los datos muestrales para saber si se relacionan y cómo se relacionan entre sí dos variables en una población.

En cualquier campo de estudio es importante pronosticar o predecir el valor de una variable de algún proceso a partir de otras variables conocidas y que están relacionadas.

Por ejemplo, un fabricante de automóviles podría estimar la demanda semanal y para un auto nuevo, basándose en una muestra aleatoria independiente de las demandas semanales; el gerente de una planta se podría interesar en conocer la relación entre el rendimiento en la obtención de un cierto producto químico y una serie de variables asociadas a este proceso de elaboración; al director de personal de una corporación, al igual que el director de admisiones en una universidad, les gustaría probar y medir algunas características individuales para que la corporación pueda contratar a la persona más adecuada para un trabajo en particular. Todos estos problemas y muchos otros problemas de pronósticos son iguales en muchos aspectos y son tratados a través de las técnicas de regresión y correlación.

El análisis de regresión da lugar a una ecuación matemática que describe la relación entre dos variables y el de correlación produce un número que resume el grado de relación entre las dos variables.

4.2. Ecuación de regresión lineal

La regresión lineal simple comprende el intento de encontrar una línea recta o ecuación matemática lineal que describa la relación entre dos variables.

El primer paso que se realiza para establecer si hay regresión es determinar si existe o no relación entre las variables, elaborando una gráfica que se denomina diagrama de dispersión, la misma se representa en el plano cartesiano en el cual la variable independiente es X y la dependiente Y.

La ecuación de regresión lineal simple es:

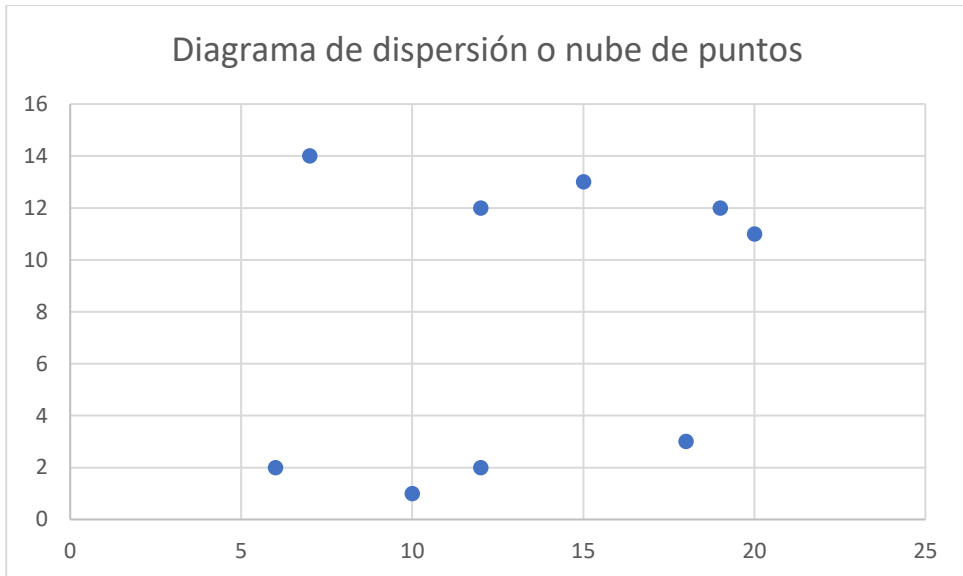
$$Y_c = a + bX.$$

4.3. Diagrama de dispersión

Es la representación gráfica más útil para describir el comportamiento conjunto de dos variables, es el diagrama de dispersión o nube de puntos, donde cada caso aparece representado como un punto en el plano definido por las variables X y Y.

A partir del diagrama de dispersión y de los resultados obtenidos en el análisis de correlación puede decidirse si esta relación es de tipo lineal. En este caso, los puntos del diagrama de dispersión aparecen tanto más próximos a una línea recta ajustada a la nube de puntos cuanto más intenso es el grado de asociación. Por otra parte, según sea el sentido de

la asociación dicha línea tendrá pendiente positiva si el coeficiente de correlación simple, r , es positivo y negativa en caso contrario.

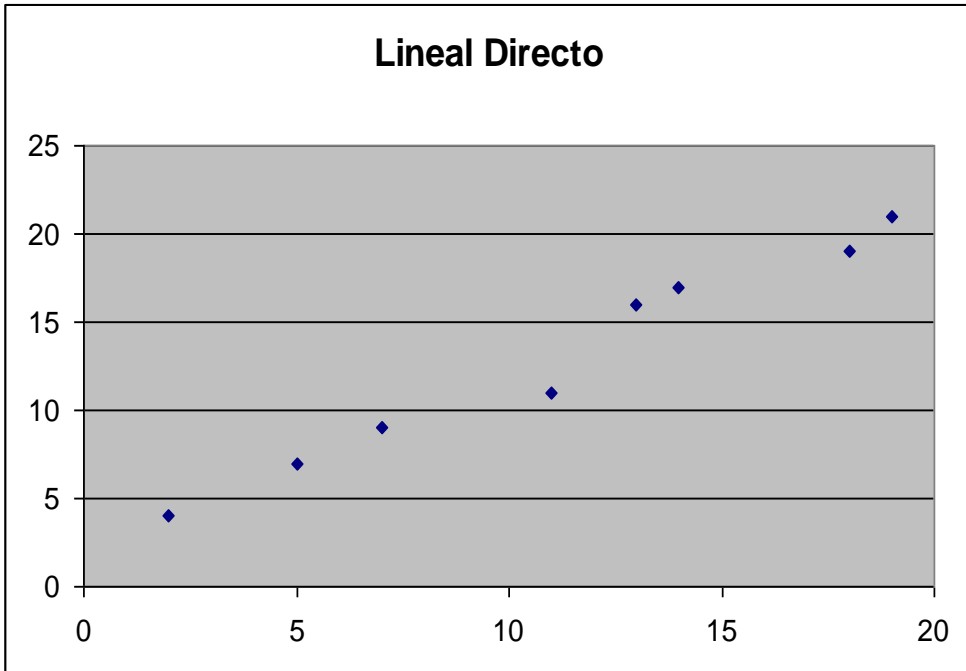


Para poder crear un modelo de regresión lineal, es necesario que se cumpla con los siguientes supuestos:

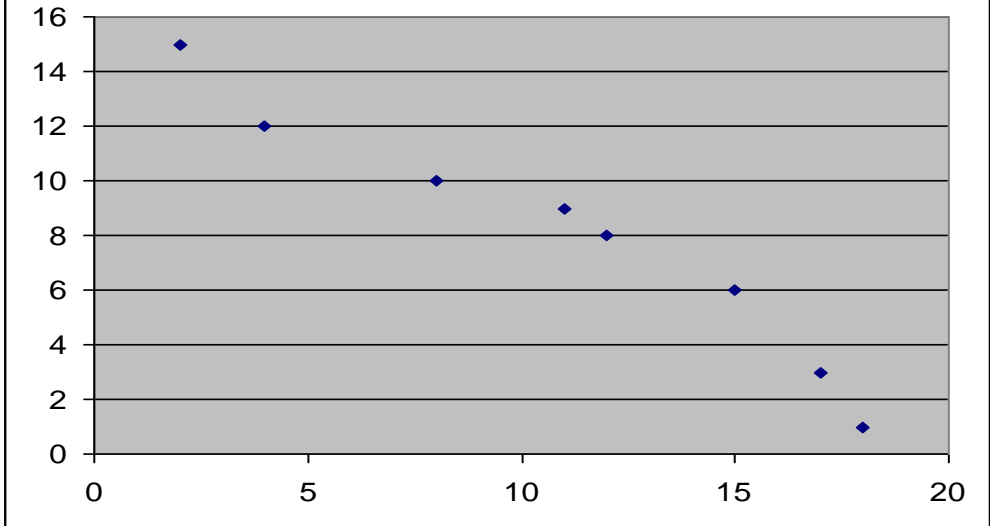
1. La relación entre las variables es lineal.
2. Los errores son independientes.
3. Los errores tienen varianza constante.
4. Los errores tienen una esperanza matemática igual a cero.
5. El error total es la suma de todos los errores.

4.4. Tipos de relación entre las variables

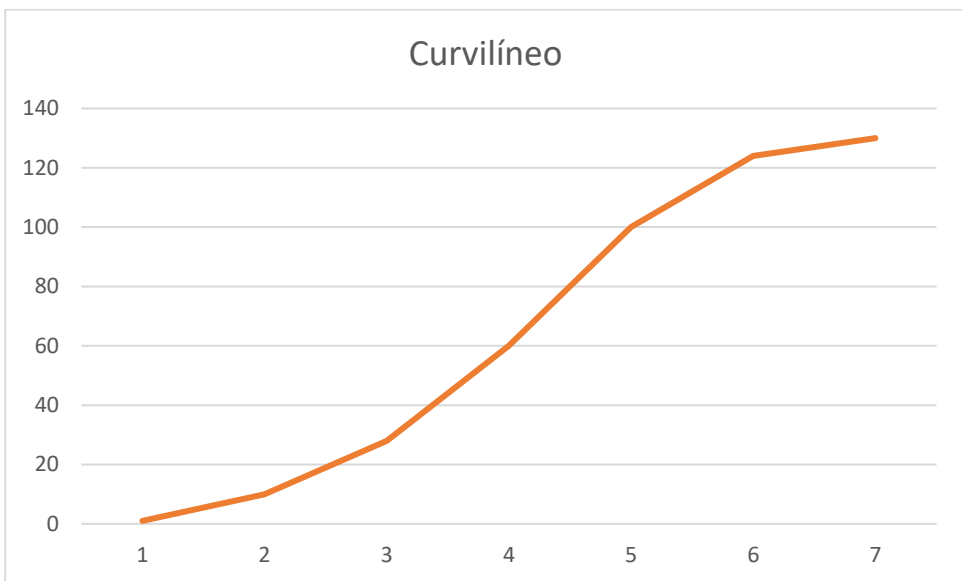
Los tipos de relación que se pueden encontrar al trazar el diagrama de dispersión son: lineal directo, lineal inverso, curvilíneo directo e inverso o sin relación.

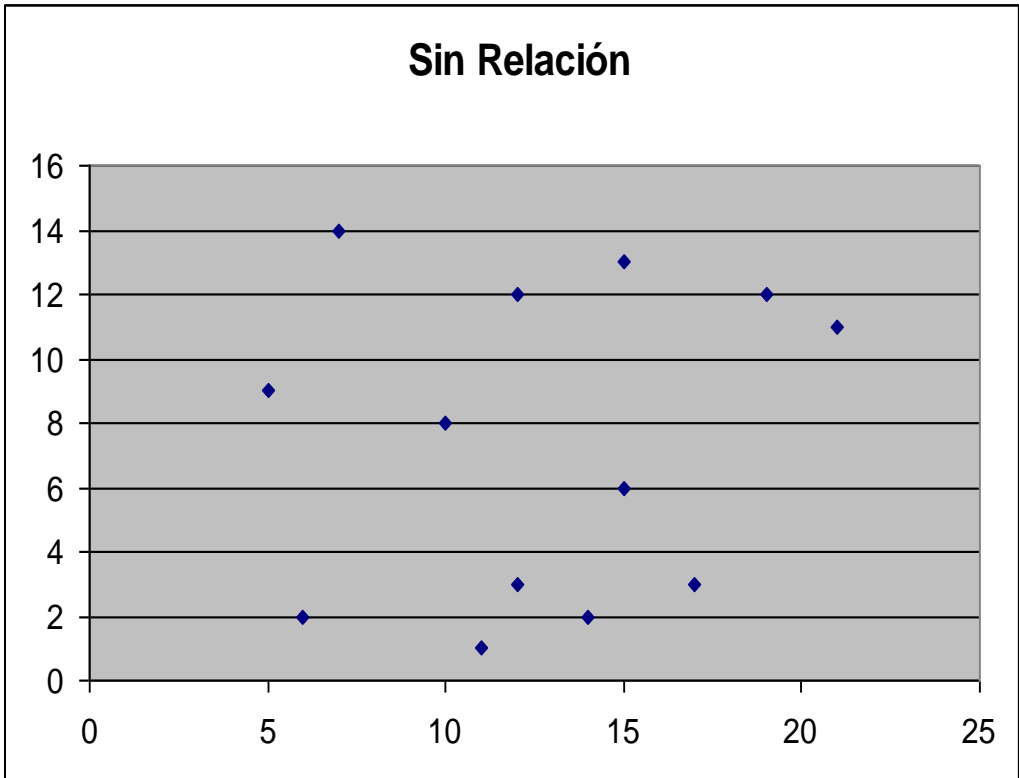


Lineal Inverso



Curvilíneo





4.5. Principio de mínimos cuadrados

El procedimiento para encontrar la recta de mejor ajuste para un conjunto de puntos, parece ser una formalización del procedimiento que se emplea para ajustar visualmente una recta. El procedimiento más utilizado para adaptar una recta a un conjunto de puntos se conoce como método de mínimos cuadrados, el cual minimiza al máximo el error de estimación; es decir, hace mínima la diferencia entre los puntos estimados y sobre la línea y los puntos reales y observados.

Para la derivación de la ecuación de regresión se parte de las ecuaciones normales:

$$\sum Y = aN + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2 \text{ donde se tiene que:}$$

$$a = \frac{\sum Y (\sum X^2) - \sum X (\sum XY)}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{N\sum XY - \sum X (\sum Y)}{N\sum X^2 - (\sum X)^2}$$

Estas fórmulas para los coeficientes de a y b se reemplazan en la ecuación de regresión lineal $Y_c = a + bx$.

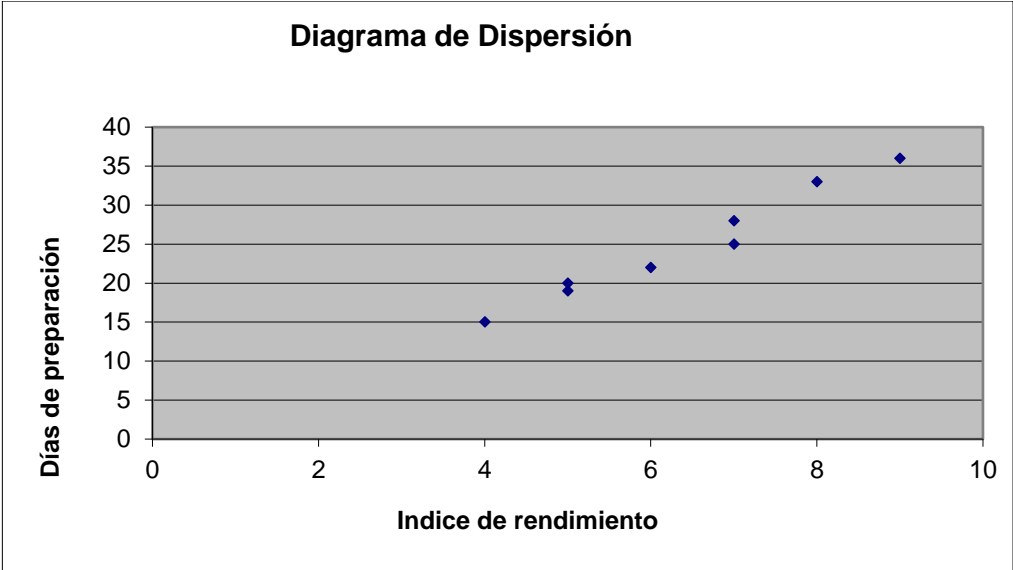
Ejemplo:

Un entrenador quiere averiguar la relación entre el índice de rendimiento en una disciplina en función a los días de preparación de los participantes. Se escogió una muestra de 8 jugadores y se reunió la información sobre los días de preparación y el rendimiento en la disciplina (1-10; 10 es el óptimo)

Índice de rendimiento

en la disciplina (X)	9	7	8	4	7	5	5	6
Días de Preparación (Y)	36	25	33	15	28	19	20	22

a- Trace el diagrama de dispersión



b- Desarrolle la ecuación de regresión

x	y	xy	x ²	Y ²	
9	36	324	81	1296	
7	25	175	49	625	
8	33	264	64	1089	
4	15	60	16	225	
7	28	196	49	784	
5	19	95	25	361	
5	20	100	25	400	
6	22	132	36	484	
$\Sigma x = 51$		$\Sigma y = 198$	$\Sigma xy = 1346$	$\Sigma x^2 = 345$	$\Sigma y^2 = 5264$

$$a = \frac{\Sigma y (\Sigma x^2) - \Sigma x (\Sigma xy)}{N \Sigma x^2 - (\Sigma x)^2} = \frac{198(345) - 51(1346)}{8(345) - (2601)} = \frac{68310 - 68616}{2760 - 2601} = \frac{-306}{159} = -2.11$$

$$b = \frac{N \Sigma xy - \Sigma x (\Sigma y)}{N \Sigma x^2 - (\Sigma x)^2} = \frac{8(1346) - 51(198)}{8(345) - (2601)} = \frac{10768 - 10098}{2760 - 2601} = \frac{670}{159} = 4.21$$

Luego la ecuación lineal resultante que predice el índice de rendimiento en función de los días de preparación resulta ser:

$$Y_c = -2.11 + 4.21x$$

4.6. Correlación lineal

Muchas relaciones son estadísticamente significativas; pocas expresan una correlación perfecta o exacta. El análisis de correlación es la herramienta estadística de la que nos valemos para describir el grado de

relación o asociación que hay entre dos variables. Las correlaciones varían de acuerdo con su dirección y a la fuerza.

El objetivo de la Correlación lineal es determinar la consistencia de una relación lineal entre las observaciones. La forma más común del análisis de correlación comprende datos continuos.

Los análisis de correlación son herramientas estadísticas que, si se emplean bien, pueden ayudar mucho en la toma de decisiones.

La medida de correlación lineal comúnmente utilizada es el coeficiente de correlación de Pearson.

La fórmula del coeficiente de correlación es la siguiente:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}}$$

Este coeficiente de correlación es útil solamente para detectar una correlación lineal entre X y Y y presenta dos propiedades que establecen la naturaleza de una relación entre dos variables. Una es su signo (+ ó -) y la otra es su magnitud.

El coeficiente de correlación producto-momento es un valor numérico que indica el grado de relación lineal entre dos variables y se le puede dar la siguiente interpretación:

$r = 1$ Correlación perfecta

$r = 0.90$ o más Correlación excelente

$0.80 < r < 0.90$ Correlación aceptable

$0.60 < r < 0.80$ Correlación regular

$0.30 < r < .0.60$ Correlación mínima

$0 < r < 0.30$ No hay correlación lineal

$r = 0$ No existe correlación

Esta tabla se utiliza tanto para valores positivos como negativos y se le da la misma interpretación para el coeficiente de determinación.

Para el ejemplo anterior de regresión lineal se tiene el coeficiente de correlación resultante que es:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{670}{\sqrt{(159) [(8)5264 - (198)^2]}}$$

$$r = \frac{670}{\sqrt{(159)(10908)}} = \frac{670}{680} = 0.985$$

Práctica de regresión y correlación

1- Se realizó un estudio para determinar la relación entre el número de años de experiencia y el salario anual para los entrenadores deportivos en particular en una región geográfica.

Salario Anual	26.1	33.2	36.1	16.5	26.4	36.4
Años de Experiencia	13	16	30	2	8	31

- Qué tipo de relación existe
- Encuentre la ecuación de regresión lineal
- Estime el salario promedio para un entrenador que tiene 12 años de experiencia
- Calcule el coeficiente de correlación

2- Los siguientes datos de la altura y el peso de una muestra de 5 jugadoras de voleibol del sexo femenino de una gran empresa.

Altura	68	67	65	68	64	(pulgadas)
Peso	119	118	129	135	123	(Libras)

- Encuentre la ecuación de regresión
- Estime la altura cuando una persona pesa 150 libras
- Calcule el coeficiente de correlación

3 - Con los ingresos y gastos de una familia deportista encuentre la ecuación de predicción y calcule el ingreso de una familia deportista que tiene 15 en gastos. (los valores son en miles)

Gastos	6.3	7.0	8.1	9.0	7.3
Ingresos	6.5	8.0	10.2	11.0	7.5

Qué tipo de relación existe entre las variables

Encuentre el coeficiente de Pearson.

4- Los siguientes son los niveles de sueldo y años de educación de 7 empleados que han laborado en la compañía de deportes por cinco años.

Años de Educación	16	12	10	8	19	18	14	13
Sueldo	2500	2230	1500	1370	5400	3400	2500	2400

- Encuentre la recta de regresión y el coeficiente de correlación
- Si una persona tiene 9 años de educación cuál será el sueldo esperado.

5- En la tabla siguiente se indica la edad (en años) y el puntaje obtenido en la ejecución de una destreza gimnástica (medida en una escala de cero a 10) de 10 niños. Edad puntaje obtenido en la ejecución de una destreza gimnástica

EDAD (X)	PUNTAJE OBTENIDO (Y)
6	9
6	6
6.7	7
7	8
7.9	9
8	9
8.2	8
8.5	8
8.9	9

Realiza lo siguiente:

- Obtener la recta de regresión del puntaje obtenido en la ejecución de la destreza gimnástica en función de la edad.
- A partir de dicha recta, obtener del puntaje obtenido en la ejecución de la destreza que correspondería a un niño de **7.5** años.

Introducción a la probabilidad y al muestreo

5.1. Conceptos básicos de probabilidad

Se origina en los juegos de Azar en el siglo XVI, se aplicó con éxito en las mesas de juego y eventualmente en los negocios.

En la actualidad, la probabilidad es parte diaria de nuestro vivir, ya sea en decisiones de carácter personal o empresarial, debido a que a través de ella podemos desarrollar estrategias donde el punto central es la capacidad de cuantificar cuán probable es determinado evento.

Experimento aleatorio

Es cualquier actividad capaz de repetirse en condiciones esencialmente estables y cuyo resultado no se puede predecir.

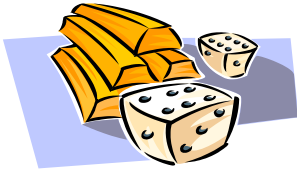
Se denominan **experimentos deterministas** aquellos que, realizados de una misma forma y con las mismas condiciones iniciales, ofrecen siempre el mismo resultado

Espacio muestral de un experimento aleatorio

Podemos definir un espacio muestral como el conjunto de todos los resultados de un experimento aleatorio, el cual se designa siempre con la letra mayúscula **S**.

Espacio muestral: conjunto de posibles resultados; punto muestral: elemento del espacio muestral y suceso: cualquier subconjunto del espacio muestral.

Ejemplo: si un experimento E, se define como lanzar un dado;



Entonces el espacio muestral asociado a tal experimento sería todos los resultados posibles donde tendremos que $S = \{1, 2, 3, 4, 5, 6\}$

Puntos muestrales

Punto muestral es cada uno de los elementos de un espacio muestral.

Al lanzar un dado los puntos muestrales serían: $\{1, 2, 3, 4, 5, 6\}$

Enfoques de Probabilidad

Teoría clásica

La definición clásica de probabilidad fue dada por Laplace y dice que la probabilidad es la razón del número de casos favorables al número total de casos igualmente posibles.

La fórmula para el cálculo de la probabilidad de un evento es la siguiente:

$$P(A) = \frac{\text{Número de elementos favorables a A}}{\text{Número total de puntos muestrales}}$$

Ejemplo 1

La probabilidad de obtener el número 4 al lanzar al aire un dado normal sería:

Casos favorables = 1 (o sea que salgan cuatro puntos del dado).

Número total de casos = 6 (total de caras del dado)

Probabilidad = $1/6$

Ejemplo 2 Probabilidad de que al lanzar un dado salga un número par:

En este caso los casos favorables son tres (que salga el dos, el cuatro o el seis), mientras que los casos posibles siguen siendo seis.

Por lo tanto: $P(A) = 3 / 6 = 0,50$ (o lo que es lo mismo, 50%)

5.2. Muestreo

El muestreo es un elemento clave en la metodología de la investigación, ya que implica seleccionar a un grupo de elementos que se utilizarán para dirigir un estudio. Es importante diseñar un plan de muestreo que defina el proceso de selección del grupo de elementos seleccionados (Espinoza,

2018). Además, de considerar la población de estudio y el enfoque del estudio.

Ahora te preguntaría ¿Cuál es el mejor tipo de muestreo, aleatorio, de conveniencia, por racimos, etc., que puedes utilizar en una investigación?

Para contestar a tu pregunta y de acuerdo con los autores consultados lo primero que tienes que considerar es que:

Fuentelzaz (2004), señala que la muestra es: “*el grupo de individuos que realmente se estudiarán, es un subconjunto de la población*” (p. 1). Es una colección de individuos extraídos de la población a partir de algún procedimiento específico para su estudio o medición directa. Una muestra es una fracción o segmento de una totalidad que constituye la población. La muestra es en cierta manera una réplica en miniatura de la población. Se estudian las muestras para describir a las poblaciones, ya que el estudio de muestras es más sencillo que el de la población completa, porque implica menor costo y demanda menos tiempo (López-Martín, et al., 2016).

Para que una muestra sea representativa, tiene que contener las características relevantes de la población en las mismas proporciones en que están incluidas en tal población. Para evaluar la representatividad de la muestra se compara la media muestral con la media poblacional, si este parámetro se desconoce se puede estimar tratando de encontrar las medias obtenidas en trabajos anteriores que han analizado las mismas variables. Siempre hay un grado de error en las estimaciones lo que constituye el error muestral, una de las maneras de minimizarlo es diseñando un plan de muestreo adecuado.

5.3. Conceptos básicos en muestreo

Población: es todo conjunto de elementos, finito o infinito, definido por una o más características, de las que gozan todos los elementos que lo componen, y sólo ellos.

Censo: en ocasiones resulta posible estudiar cada uno de los elementos que componen la población, realizándose lo que se denomina un censo, es decir, el estudio de todos los elementos que componen la población.

Muestra: una muestra es un subconjunto de elementos de una población. Para extraer conclusiones válidas e imparciales referidas a todos los elementos de la población a partir de la observación de sólo unos pocos elementos, es necesario, que la muestra utilizada sea representativa de la población; esto se consigue mediante las “técnicas de muestreo”.

Unidades de muestreo: número de elementos de la población, no solapados, que se van a estudiar. Todo miembro de la población pertenecerá a una y sólo una unidad de muestreo.

Unidad estadística o de análisis.

Se llama unidad estadística o individuo a cada uno de los elementos que componen la población estadística. El individuo es un ente observable que no tiene por qué ser una persona, puede ser un objeto, un ser vivo o incluso algo abstracto.

De acuerdo a Hernández-Sampiere, Fernández-Collado, Baptista-Lucio (2014), definen la unidad de análisis como la “*entidad principal que se está analizando en un estudio. Es el "qué" se está estudiando o a "quién" se está estudiando*” (p. 173). En la investigación en Ciencias Sociales,

las unidades típicas de análisis incluyen individuos (más comunes), grupos, organizaciones sociales y artefactos sociales.

Este concepto no debe confundirse con la unidad de observación, que es la unidad descrita por los datos de uno (vecindarios que usan el Censo de los EE. UU., individuos que usan encuestas, etc.). Por ejemplo, un estudio puede tener una unidad de observación a nivel individual, pero puede tener la unidad de análisis a nivel de vecindario, extrayendo conclusiones sobre las características de vecindario a partir de datos recopilados de individuos.

Arias-Gómez et al., (2016), sostienen que la población estadística es una variable aleatoria relacionada con los objetos o individuos que se pretende estudiar en una investigación. Cada uno de los elementos de la población se denomina individuo y estos comparten algunas características como:

- Homogeneidad. Todos los miembros de la población tienen las mismas características según las variables a estudiar.
- Temporalidad. El periodo donde se sitúa a la población objeto de estudio. Debe establecerse si el estudio es presente o retrospectivo.

Marco muestral: lista de unidades o elementos de muestreo.

Tipos de muestreo.

Fuentelaz (2004), señala que las muestras pueden ser:

- Muestras aleatorias o probabilísticas en cada uno de los elementos de la población tiene la misma probabilidad de integrar parte de la

muestra. Las unidades de análisis tienen que estar dispuestas de modo tal que el proceso de selección dé una equis (X) probabilidad de selección a todas y cada una de las unidades que figuran en la población. Cada uno de los elementos de la población tiene la misma probabilidad de integrar parte de la muestra (Otzen y Manterola, 2017).

- Muestras no probabilísticas nos dicen que la elección de los elementos no depende de la probabilidad, sino de las causas relacionadas con las características de la investigación o de la muestra. La elección de los elementos no depende de la probabilidad, sino de las causas relacionadas con las características de la investigación o de quién establece la muestra.
- Aquí el procedimiento no es mecánico ni en base a fórmulas de probabilidad, sino que depende del proceso de toma de decisiones de una persona o de un grupo de personas y, desde luego, las muestras seleccionadas obedecen a otros criterios de investigación. Desde el enfoque cuantitativo y para determinado diseño, la utilidad de una muestra no probabilística reside no tanto en una “*representatividad*” de elementos, sino en una cuidadosa y controlada elección de sujetos con ciertas características definidas previamente en el planteamiento del problema. (Otzen y Manterola, 2017).

5.4 Técnicas de muestreo

De acuerdo con Otzen y Manterola (2017), la representatividad de una muestra permite extrapolar y, por lo consiguiente, generalizar los

resultados observados en esta, a la población objeto de estudio. Entonces, el muestreo tiene por objetivo estudiar las relaciones existentes entre la distribución de una variable en la población y la distribución de esta variable en la muestra a estudio. Para ello, es fundamental, entre otras cosas definir, los criterios de inclusión (características clínicas, demográficas, temporales y geográficas de los sujetos que componen la población en estudio) y de exclusión (características de los sujetos que pueden interferir con la calidad de los datos o la interpretación de los resultados), por lo que hay diferentes técnicas de muestreo y las más utilizadas en la investigación son:

- Muestreo aleatorio simple. Se utiliza cuando se conocen todos los elementos que conforman la población. Cada elemento de la población tiene la misma posibilidad de ser elegido para formar parte de la muestra. Se elabora una lista enumerativa de los elementos. Desde aquí se obtiene una serie de elementos que constituyen la muestra. El proceso consiste en trabajar con una tabla de dígitos al azar o una serie de números aleatorios generados en una planilla de cálculo y seleccionar a los que coinciden con el número al azar y el orden de la lista o utilizando un programa estadístico que seleccione una cantidad de elementos aleatoriamente de una lista general que es la población.
- Muestreo estratificado. Esta técnica de muestreo resulta apropiada cuando la población ya está dividida en grupos, porque refleja de forma más precisa las características de la población y permite efectuar comparaciones entre los estratos conformados. Como,

por ejemplo: grupos de edad, grupos étnicos, por mencionar algunos.

- Muestreo por conglomerados o racimos. Se utiliza cuando no es posible obtener una lista de todos los elementos de la población. Su empleo es adecuado si la población es muy grande y dispersa. Los conglomerados se caracterizan por ser homogéneos entre sí, pero internamente presentan un alto grado de heterogeneidad en sus componentes.
- Muestreo sistemático. Los elementos se seleccionan de la población en un intervalo uniforme que se mide respecto de tiempo, orden o espacio. Se emplea si existe una lista ordenada de los elementos de la población o cuando se sabe cuántos elementos componen esa población.

Técnicas de muestreo no probabilístico. Por cuotas, opinático o intencional, casual o incidental, por redes o bola de nieve.

- Muestreo por cuotas. Consiste en formar estratos de la población sobre la base de ciertas características y en procurar que estén representadas en proporciones semejantes a las que existen en la población. Principales características utilizadas: sexo, edad, ocupación, etc.

Una vez determinada la cuota se eligen los primeros que se encuentran y que cumplen esas características. Este tipo de muestreo tiene como beneficio que se pueden realizar estudios exploratorios rápidos y económicos.

- Muestreo opinático o intencional. Es una técnica que se basa en la opinión del investigador para constituir una muestra de sujetos en función de su carácter típico, como en el estudio de casos extremos o marginales, o de los casos típicos. Permite estudiar fenómenos raros o inusitados.

Ejemplo:

Encuesta sobre autocuidado a personas que han recibido trasplante hepático en un hospital determinado.

- Muestreo casual o incidental. La muestra está conformada por sujetos fácilmente accesibles y presentes en un lugar determinado, y en un momento preciso. Los sujetos se incluyen en el estudio a medida que se presentan, y hasta que la muestra alcance el tamaño deseado.

Ejemplo: encuestas en vía pública que se realizan en un día y horario determinado.

- Muestreo por redes (bola de nieve).

Consiste en localizar a algunos individuos según determinadas características. Se utiliza en poblaciones marginales o de difícil acceso. Se basa en redes sociales, en las amistades. Cuando se encontró el primer representante, este puede conducir a otro, y ese a un tercero, y así sucesivamente hasta conseguir una muestra suficiente.

Ejemplos: Consumidores de productos premium. Personas que padecen determinada enfermedad. Miembros de una secta. Adictos

que rechazan la instancia de rehabilitación. Mujeres golpeadas.
Aficionado a prácticas culturales inusuales.

Fórmulas para el tamaño de muestra

Muestreo aleatorio simple

FÓRMULA

$$n = \frac{N Z^2 p q}{e^2(N-1) + Z^2 p q}$$

Si se conoce el tamaño de la población:
N = tamaño de la población.

n = tamaño necesario de la muestra.

Z = margen de confiabilidad o número de unidades de desviación estándar en la distribución normal que producirá un nivel deseado de confianza.

p y q = Cuando no se tiene idea clara de esta situación, es necesario dar sus máximos valores, tanto a la probabilidad de que se realice el evento favorable, como a la de que no se realice: Esto es 50% a p y 50% a q.

E = error o diferencia máxima entre la media muestral y la media de la población que se está dispuesto a aceptar con un nivel de confianza que se ha definido.

Si no se conoce el tamaño de la población:

FÓRMULA

$$n = \frac{Z^2 p q}{e^2}$$

n = tamaño necesario de la muestra.

Z = margen de confiabilidad o número de unidades de desviación estándar en la distribución normal que producirá un nivel deseado de confianza.

p y q = Cuando no se tiene idea clara de esta situación, es necesario dar sus máximos valores, tanto a la probabilidad de que se realice el evento favorable, como a la de que no se realice:

E = error o diferencia máxima entre la media muestral y la media de la población que se está dispuesto a aceptar con un nivel de confianza que se ha definido.

La investigación constituye una actividad fundamental para los profesionales. Para evitar los sesgos de selección que son los errores referidos al proceso de identificación de la población a estudiar y selección de un espacio muestral donde se realizará el estudio, es necesario recurrir a las técnicas de muestreo para fundamentar y ofrecer verificabilidad al estudio realizado.

Así que elegir una muestra probabilística o no probabilística depende de los objetivos del estudio, del esquema de investigación y del alcance de sus contribuciones.

Es fundamental expresar claramente que en todo trabajo de investigación los aspectos del diseño metodológico: la población y las características que deben poseer los elementos para formar parte del estudio, el número de elementos que conforman la población, si este número se conoce con certeza o se puede estimar, el tipo de muestreo y la técnica utilizada, así como el tamaño de la muestra y el error de muestreo establecido serán lo más apegados con los resultados esperados.

Cómo ya te has percatado una muestra, es un subconjunto de una población. Las muestras representativas de una población son útiles, ya que facilitan el manejo de los datos. Una muestra es representativa de la población si al escogerla cada elemento tiene la misma probabilidad de salir o de ser escogido. Para que se puedan generalizar a la población los resultados obtenidos en la muestra, esta ha de ser representativa de dicha población. Para ello, se han de definir con claridad los criterios de inclusión y exclusión y, sobre todo, se han de utilizar las técnicas de muestreo apropiadas para garantizar dicha representatividad (Fuentelaz, 2004).

La teoría del muestreo permite determinar de manera efectiva la muestra que refleje con exactitud las características de la población sometida al estudio, ya que no siempre es posible tomar a todos los elementos que conforman la población. El tamaño de la muestra a estudiar y la forma de selección de las unidades muestrales se debe justificar convenientemente conforme al planteo del problema, la población, los objetivos y el propósito de la investigación (López-Martín, et al., 2016).

5.5. Selección y tamaño de muestra

La tabla de dígitos aleatorios: para utilizar una tabla de dígitos aleatorios deben seguirse los siguientes pasos:

- 1- Hacer una lista de los elementos de la población.
- 2- Numerar consecutivamente los elementos de la lista, empezando por el cero (0, 00, 000.etc.).
- 3- Tome los dígitos de una tabla de números aleatorios, de manera que la cantidad de dígitos seleccionados sea igual a la del último elemento numerado de su lista.
- 4- Omitir los dígitos que no correspondan con los números de la lista o que repita cifras seleccionadas anteriormente.
- 5- Utilizar los grupos de dígitos aleatorios seleccionados para identificar los elementos de la lista que se habrán de incluir en la muestra.

Cálculo del tamaño de muestra

- 1- Se determina el grado de confianza con el que se va a trabajar. Generalmente, se trabaja con un grado de confianza del 95%, por lo que el valor de z corresponderá a 1.96, valor establecido en la tabla de la distribución normal.
- 2- Evaluar la probabilidad de que se realice el evento favorable:

Cuando no se tiene idea clara de esta situación, es necesario dar sus máximos valores, tanto a la probabilidad de que se realice el evento favorable, como a la de que no se realice: Esto es 50% a p y 50% a q.

3- Se determina el error máximo que puede ser aceptado en los resultados. Se fija en un 5%, ya que, generalmente, se trabaja en la práctica con un grado de precisión entre el 2 y 6% para un 95% de confianza.

4- Determinación del tamaño de muestra

Fórmula para poblaciones finitas:

$$n = \frac{N Z^2 p q}{e^2(N-1) + Z^2 p q}$$

Donde:

Z = Nivel de Confianza; N= Población Total

p= Probabilidad a favor; q= Probabilidad en contra;

e= Error de estimación

Fórmula para poblaciones infinitas:

$$n = \frac{Z^2 p q}{e^2}$$

Ejemplo 1:

Suponga que va a realizar un diagnóstico sobre la competitividad de las empresas que venden insumos deportivos: Estas están clasificadas en

grandes, medianas y pequeñas empresas: 2180 son pequeñas empresas, 980 son medianas empresas y 740 grandes empresas.

¿Si en total hay 3900 empresas que venden insumos deportivos de qué tamaño debe ser la muestra de empresas por diagnosticar?

Solución:

$$n = \frac{(1.96)^2(3900)(0.5)(0.5)}{(.05)^2(3899) + (1.96)^2(0.5)(0.5)} = \frac{3745.56}{9.75 + 0.9604} = \frac{3745.56}{10.71}$$

$n = 350.$

Muestreo estratificado

En un muestreo aleatorio estratificado la población se divide en estratos o subgrupos significativos y se obtiene una muestra aleatoria de cada uno. El razonamiento es que, mediante el ordenamiento de los elementos de la población en subgrupos homogéneos, la variabilidad es menor que la de la población total, y por ello se necesita un tamaño de muestra más pequeño.

Es importante estratificar una población usando variables relevantes al estudio.

El muestreo estratificado se usa cuando cada grupo presenta una variación pequeña en su interior, pero existe una amplia variación entre ellos.

Ejemplo :

Suponga que va a realizar un diagnóstico sobre la competitividad de las empresas que venden insumos deportivos: Estas están clasificadas en

grandes, medianas y pequeñas empresas: 2180 son pequeñas empresas, 980 son medianas empresas y 740 grandes empresas.

¿Si en total hay 3900 empresas que venden insumos deportivos de qué tamaño debe ser la muestra de empresas por diagnosticar? ¿Si se realiza en proporción al tamaño de cada sector cuánto sería el tamaño de muestra indicado?

Solución:

Se tiene que $n = 350$ es el tamaño de muestral de empresas en general
Luego el tamaño correspondiente para cada estrato según el tamaño de la empresa sería:

$$\text{Empresas Pequeñas } n_{h1} = \frac{2180}{3900} (350) = 196$$

$$\text{Empresas Medianas } n_{h2} = \frac{980}{3900} (350) = 88$$

$$\text{Empresas Grandes } n_{h3} = \frac{740}{3900} (350) = 66$$

Práctica de muestreo

1- Se realiza un estudio en la ciudad de Penonomé a 150 familias de clase media, para conocer el tipo de deporte preferido que ellos ven por televisión en su hogar ¿Cuál es la población?, ¿Cuál es la muestra?, ¿Cuál es la unidad?, ¿Cuál es la característica? ¿La característica en estudio es cuantitativa o cualitativa?

2- Se desea realizar un estudio sobre las condiciones socioeconómicas de los residentes del distrito de Portobelo en la provincia de Colón. Según el censo de población del año 2000 el número de viviendas ocupadas fue de 1902. Si se estima un error del 5% y una confianza del 90%, encuentre el número de viviendas necesarias que hay que encuestar. Se desea realizar el estudio por el método de muestreo estratificado.

Distrito: Portobelo	Corregimiento	N° de Viviendas
	Portobelo	942
	Cacique	60
	Garrote	181
	Isla Grande	237
	María Chiquita	482

3- El departamento de personal de una empresa de venta de equipos deportivos para una nueva ubicación de sus empleados, decide llevar a cabo un estudio sobre la situación económica de aquellos.

a)- ¿Cuál es la población de interés?

b)- ¿Cuáles son las características que interesa estudiar?

c)- ¿Cuál es la unidad estadística?

4- Suponga que se desea efectuar un sondeo de opiniones de residentes en su barrio para conocer cuál es su deporte favorito. Describa cómo podría realizar esta tarea si se decidiera por usar: muestreo simple al azar y por muestreo estratificado.

5- A continuación, se presentan ejemplos de población escriba al lado si la población es finita o infinita.

a- Salarios de los futbolistas durante al año anterior. _____

b- Personas a las que se le puede aplicar una vacuna _____

c- Estudiantes que llegaran a participar de los diversos clubes deportivos en los próximos años _____

d- Resultados obtenidos al lanzar sin límite un dado _____

e- Producción de energía eléctrica en los últimos 10 años _____

REFERENCIAS

- Abello Llanos, R. (2010). La investigación en ciencias sociales: sugerencias prácticas sobre el proceso. *Revista Investigación y Desarrollo*, 17(1), 208-26.
- Aignerren, M. (2010). Técnicas de medición por medio de escalas. Centro de Estudios de Opinión de la Universidad de Antioquía. Recuperado de: <file:///D:/Downloads/6552-18195-1-PB.pdf>
- Alarcón, A. M. y Muñoz, S. (2008). Medición en salud: algunas consideraciones metodológicas. *Revista Médica de Chile*, (138), 125-30.
- Artés Rodríguez, E. (2006). Estadística aplicada al deporte. Universidad de Almería, España. Recuperado de: <file:///D:/Documents/INVESTIGACIONES/Supervisi%C3%B3n%20en%20educ%20física/23160.pdf>
- Badii, M. H., Castillo, J., Landeros, J. y Cortes, K. (2007). Papel de la estadística en la investigación científica. *Revista de Innovaciones de Negocios*, 4(1), 107-145.
- Bar, A. (2000). Un aporte a la discusión sobre el status metodológico de las variables y escalas de medición. *Revista Cinta Moebio*, (7), 121-26. Recuperado de: www.moebio.uchile.cl/07/bar01.htm
- Campo Aranda, T. J. y Gómez Araújo, E. (2009). Técnicas e instrumentos cualitativos de recogida de datos. Recuperado de: mestrado.prpg.ufg.br/up/97/o/Técnicas_e_instrum._cualitat.Libro.pdf
- Carmona, T. G. (1997). Métodos y técnicas cualitativas de investigación en las ciencias sociales. *Revista Trabajo*, (3), 163-8.
- Cazau, P. (2006). Introducción a la investigación en ciencias sociales. 3ª edición. Editorial Galeón. Buenos Aires.
- Coronado Padilla, J. (2007). Escalas de medición. *Revista Paradigmas*, 2(2), 104-25.

- Dávila Newman, G. (2006). El razonamiento inductivo y deductivo dentro del proceso investigativo en ciencias experimentales y sociales. *Revista Laurus*, 12, 180-205.
- Del Canto, E. y Silva Silva, A. (2013). Metodología cuantitativa: abordaje desde la complementariedad en ciencias sociales. *Revista Ciencias Sociales*, 141(3), 25-4.
- Díaz-Levicoy, D., Batanero, C., Arteaga, P. y López-Martín, M. del M. (2015). Análisis de los gráficos presentados en libros de texto de educación primaria chilena. *Revista de Educación Matemática*, 16(4), 715-39.
- Espinoza Salvadó, I. (2016, marzo). *Tipos de muestreo*. [Diapositivas de PowerPoint]. Recuperado de: <http://www.bvs.hn/Honduras/Embarazo/Tipos.de.Muestreo.Marzo.2016.pdf>
- Fuentelsaz Gallego, C. (2004), Cálculo del tamaño de la muestra. *Revista Matronas de Profesión*, 5(18), 1-9.
- Ferreiro, O. y Fernández de la R., P. (1998). La estadística, una ciencia en la controversia. *Revista Universitaria*, (25). Recuperado en: <http://dta.utalca.cl/estadistica/documentos/articulo4.pdf>
- Gamboa Graus, M. E. (2018). Estadística aplicada a la investigación educativa. *Revista Dilemas Contemporáneos: Educación, Política y Valores*, 5(2), 1-32. Recuperado en: <http://www.dilemascontemporaneoseducacionpoliticayvalores.com>
- Gil Flores, J. (2003). La estadística en la investigación educativa. *Revista de Investigación Educativa*, 21(1), 231-48.
- Lafuente Ibáñez, C. y Marín Egoscozabal, A. (2008). Metodología de la investigación en las ciencias sociales. *Revista EAN*, (64), 5-18.
- López-Martín, M del M., Batanero, C., Díaz-Batanero, C. y Gea, M. M. (2016). La inferencia estadística en las pruebas de acceso a la Universidad de Andalucía. *Revista Paranense de Educación Matemática*, 5(8), 33-59.

- Manterola, C. y Otzen, T. (2015). Los sesgos de la investigación clínica. *Revista Temuco*, 33(3), 1156-1164.
- Martínez, Ciro.(2005). Estadística y Muestreo. ECOE EDICIONES.
- Mejía Navarrete, J. (2003). Técnicas cualitativas de investigación en las ciencias sociales. *Revista Investigaciones Sociales*, (8), 223-56.
- Otzen, T. y Manterola, C. (2017). Técnicas de muestreo sobre una población a estudio. *Revista Temuco*, 35(1), 227-232.
- Piña León, L. y de Rojas Gómez, H. (2018). Estadística aplicada a la actividad empresarial (I) Técnicas de muestreo y auditoría (I). *Revista Economía y Desarrollo*, 145(1 y 2), 197-220.
- Rivera, J. H. (2017). El estado de la investigación en Colombia. *Revista Avenir*, 1(2), 1-4.
- Romo M., M. y Castillo D., C. (2002). Metodología de las ciencias sociales aplicadas al estudio de la nutrición. *Revista Chilena de Nutrición*, (29). Recuperado de:
scielo.conicyt.cl/scielo.php?pid=S0717-75182002000100003&script=sci_arttext
- Triola, M. (2009) Estadística. 10ª Edición; México, Pearson Educación. [versión virtual]