BIOESTADÍSTICA EDUCATIVA

Prof. Joan Fernando Chipia Lobo

Título de la obra: Bioestadística Educativa

Autor: Joan Fernando Chipia Lobo.

Primera Edición: 2015.
Segunda Edición: 2016.
Tercera Edición: 2018.
Mérida, Venezuela.

TABLA DE CONTENIDO	Pág.

INTRODUCCIÓN

Las actitudes populares hacia la Estadística están impregnadas por una mezcla de terror, de cinismo, de recelo y de desprecio. Sin ningún miramiento, los estadísticos han sido tachados de mentirosos y se les ha acusado del delito de ejercer la *estadistificación*, el arte de mentir con estadísticas, conservando una apariencia de objetividad y veracidad.

Alguien dijo que "si todos los estadísticos del mundo se pudieran callar de una vez por todas, mucho se ganaría". Al estadístico se le describe desdeñosamente como alguien que se ahoga en un vaso de agua o como aquél cuya cabeza está en la nevera y cuyos pies están en el horno y que dice que en general se siente bien.

El escritor irlandés George Bernard Shaw (1856-1950), señaló que: *"la estadística es una ciencia que demuestra que si mi vecino tiene dos coches y yo ninguno, los dos tenemos uno"*. Aquellos que estudian por primera vez *estadística,* harán bien en desechar la imagen popular de esta ciencia, así como la de los estadísticos, puesto que deben darse cuenta de que resulta igualmente fácil decir despropósitos como expresarlos cuantitativamente.

Por lo tanto, una formación lógica permite identificar los absurdos verbales, y por su parte, un buen conocimiento en Estadística es la mejor arma para eliminar los absurdos cuantitativos. Entonces para eliminar la imagen popular de estadística, es necesario estudiar: sus métodos y sus características históricas.

El libro se divide en: Capítulo 1. Breve Historia de la Estadística y Bioestadística; Capítulo 2. Introducción a la Bioestadística; Capítulo 3. Conceptos Básicos de Bioestadística; Capítulo 4. Representación y Organización de datos; Capítulo 5. Medidas de Tendencia Central y de Variabilidad; Capítulo 6. Medidas de Posición; Capítulo 7. Probabilidad elemental; Capítulo 8. Distribuciones de Probabilidad; Capítulo 9. Pruebas de hipótesis Estadísticas y Capítulo 10. Muestreo.

CAPÍTULO 1. BREVE HISTORIA DE LA ESTADÍSTICA Y BIOESTADÍSTICA.

HISTORIA DE LA ESTADÍSTICA

Desde los comienzos de la civilización han existido formas sencillas de estadísticas, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales y otras cosas.

Su origen empieza posiblemente en la *Isla de Cerdeña, 3500 a. C.* donde existen monumentos prehistóricos pertenecientes a la civilización *Nurágica*, los primeros habitantes de la isla; en donde los Nuragas, servían para llevar la cuenta del ganado y la caza.

En el antiguo *Egipto*, los faraones lograron recopilar, alrededor del año *3050 a. C.,* prolijos datos relativos a la población y la riqueza del país; con el propósito de preparar la construcción de las pirámides. También, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto. Hacia el año 3000 a. de C. los *Babilonios* utilizaban ya pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque. En China ya había registros numéricos, porque existían los censos chinos ordenados por el emperador Tao hacia el año *2.200 a.C.*

En el antiguo Israel, la *Biblia* da referencia, en el libro de los Números (*1440 a 1400 a. C.*), de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David, por otra parte, ordenó a Joab, general del ejército, hacer un censo de Israel con la finalidad de conocer el número de habitantes, y el libro Crónicas describe el bienestar material de las diversas tribus judías.

Los *Griegos* hacia el año *594 a. de C.*, efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

Los *Romanos (500 a. C.),* fueron maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años llevaban a cabo un censo de la población, y los funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. En la época del nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del Imperio.

En el año 762, *Carlomagno* ordenó la creación de un registro de todas sus propiedades, así como de los bienes de la iglesia. El rey *Guillermo I* encargó un censo en el año *1086* en *Inglaterra*. La información en él obtenida se recoge en el Libro del Gran Catastro, que es un documento acerca de la propiedad, la extensión y el valor de las tierras en Inglaterra. Esta obra fue el primer compendio estadístico de ese país. El rey *Guillermo I* encargó un censo en el año *1086* en *Inglaterra*. La información en él obtenida se recoge en el Libro del Gran Catastro, que es un documento acerca de la propiedad, la extensión y el valor de las tierras en Inglaterra. Esta obra fue el primer compendio estadístico de ese país.

Durante los siglos XV, XVI y XVII: Leonardo de Vinci, Nicolás Copérnico, Galileo Galilei, William Harvey, Francis Bacon y René Descartes hicieron grandes operaciones con base en el método científico, de tal forma que cuando se crearon los Estados nacionales y surgió como fuerza el comercio internacional, había ya un método capaz de aplicarse a los datos económicos.

El término alemán *STATISTIK*, que fue primeramente introducido por *Gottfried Achenwall* en *1749,* designaba originalmente el análisis de datos del Estado, es decir, la "ciencia del Estado" (también llamada aritmética política de su traducción directa del inglés). No fue hasta el siglo XIX cuando el término estadística adquirió el significado de recolectar y clasificar datos. Este **c**oncepto fue introducido por el militar británico Sir John Sinclair (1754-1835).

HISTORIA DE LA BIOESTADÍSTICA

El primer médico que utilizó métodos matemáticos para cuantificar variables de pacientes y sus enfermedades fue el francés Pierre Charles-Alexandre Louis (1787-1872), en un estudio de la tuberculosis, que influyó en toda una generación de estudiantes. Sus discípulos, a su vez, reforzaron la nueva ciencia de la epidemiología con el método estadístico.

Otros eventos a considerar son los de William Heaton Hamer (1862-1936) propuso un modelo temporal discreto en un intento de explicar la ocurrencia regular de las epidemias de sarampión.

John Brownlee (1868-1927), luchó durante veinte años con problemas de cuantificación de la infectividad epidemiológica. Ronald Ross (1857-1932) exploró la aplicación matemática de la teoría de las probabilidades con la finalidad de determinar la relación entre el número de mosquitos y la incidencia de malaria en situaciones endémicas y epidémicas. El cambio más radical en la dirección de la epidemiología se debe a Austin Bradford Hill (1897-1991) con el ensayo clínico aleatorizado y, en colaboración con Richard Doll (1912), el épico trabajo que correlacionó el tabaco y el cáncer de pulmón.

El precursor histórico y llamado padre de la Bioestadística fue el inglés: Sir Francis Galton (1822-1911), introdujo un método matemático para el ajuste de curvas a puntos experimentales: el de los mínimos cuadrados. Además formuló los conceptos de Regresión y Correlación.

Karl Pearson aplicó la estadística a los problemas biológicos de la herencia y la evolución, resaltándose la publicaciones realizadas entre 1893-1912 tituladas "Contribuciones de la Matemática a la teoría de la Evolución", en las cuales se encuentran contribuciones al Análisis de Regresión, Coeficiente de Correlación. Descubrió la Distribución Chi-cuadrada y fue el quien acuñó el término Desviación Estándar.

Ronald A. Fisher (1890-1962), considerado el creador del 50% de la Bioestadística actual. Fisher realizó muchos avances en la estadística, siendo una de sus más importantes contribuciones, la inferencia estadística creada por él en 1920. En Cambridge en 1912, estudió la teoría de errores. Se dedicó al estudio pionero de los principios del diseño de experimentos (*The Design of Experiments*, 1935). Elaboró sus trabajos sobre el Análisis de Varianza (procedimiento utilizado en todo el mundo).

Referencias

Hernández, S. (2005). *Historia de la estadística.* Revista de divulgación científica y tecnológica de la Universidad Veracruzana [En línea], 18, (8). Disponible: http://www.uv.mx/cienciahombre/revistae/vol18num2/articulos/historia/index.htm [Consulta: 2012 Abril 20].

CAPÍTULO 2. INTRODUCCIÓN A LA BIOESTADÍSTICA.

Constantemente buscamos información para tomar decisiones. Al levantarnos en la mañana observamos cómo está el tiempo para decidir cómo vestirnos o averiguamos si va a llover para decidir si llevamos paraguas, etc. Inconscientemente aplicamos la lógica estadística para tomar ese tipo de decisiones. Además, podemos usar la estadística para situaciones tan diversas como:

- Analizar si el tabaquismo se asocia al cáncer al pulmón.
- Analizar si la reforma educacional mejoró la calidad de la Educación en Venezuela.
- Predecir los resultados de las próximas elecciones.
- Predecir si ocurrirá una situación de emergencia ambiental en Mérida.

Aprender Estadística es como aprender un nuevo idioma.

Estadística y Bioestadística

Deriva su nombre del hecho de haber sido aplicada primeramente a la recolección de datos que permitieran la administración de los estados, con propósitos militares e impositivos, preocupación fundamental de los Imperios, debido a que los gobernantes necesitaban conocer información referente al número y riquezas de sus súbditos.

Se derivan dos significados:

1. El término *estadísticas,* en plural, es sinónimo de datos numéricos.
2. *Estadística*, en singular, es el método utilizado en el manejo de datos.

Por lo tanto la *Estadística* es un método con procedimientos lógicos que logran el máximo de la experimentación científica. A pesar de la sencilla caracterización de *Estadística*, ésta nos permite entrever vastos campos de acción, pudiéndose decir que no hay prácticamente rama del saber humano en donde no tenga utilización. Sin embargo, lo anterior no quiere decir que la Estadística, sea el único mecanismo a través del cual puedan obtenerse nuevos conocimientos o que el solo hecho de manejar una gran cantidad de datos numéricos implique un trabajo científico.

Especificando de una vez se define a la Estadística, como una ciencia que estudia la recolección, análisis e interpretación de datos, ya sea para ayudar en la toma de decisiones o para explicar condiciones regulares o irregulares de algún fenómeno, problema o estudio aplicado, de ocurrencia en forma aleatoria o condicional. Sin embargo estadística es más que eso, en otras palabras, es el vehículo que permite llevar a cabo el proceso relacionado con la investigación científica. Es una ciencia transversal a una amplia variedad de disciplinas, desde la física hasta las ciencias sociales, desde las ciencias de la salud hasta el control de calidad. Se usa para la toma de decisiones en áreas de negocios o instituciones gubernamentales.

Cuando se hace referencia a la *Bioestadística*, se considera una rama de la estadística aplicada a la ciencias de la vida, la cual ha sido clave en el desarrollo de nuevos fármacos, en el entendimiento de enfermedades crónicas; la estrecha relación de la Estadística con el método científico hace de la Bioestadística una disciplina imprescindible en la mayoría de los proyectos de biológicos y de salud.

El pensamiento estadístico aplicado a las ciencias de la salud no sólo resuelve y entiende compleja metodología para dar respuesta a hipótesis, sino que es capaz de organizar el "sistema" que involucra la investigación desde el diseño general de experimentos en el área específica, diseño de muestreo, control de calidad de la información, análisis y presentación de resultados.

Estadística Descriptiva e Inferencial

Cuando los métodos estadísticos empezaron a desarrollarse aparecieron dos *tendencias* muy divergentes. Una de estas era una tendencia un tanto *femenina* dedicada a llevar registros ordenados de datos del Estado (*estado* y *estadística* vienen de la misma raíz latina *status*). La otra parte, es *masculina* debido a que se basa en la Matemática para incrementar el conocimiento sobre probabilidades.

La parte femenina comprendía tareas como contar, medir, describir, tabular, ordenar y levantar censos, lo cual dio lugar a la *Estadística Descriptiva* moderna. La parte masculina, generó la *Estadística Inferencial* fundamentalmente basada en la teoría de probabilidades y en una derivación llamada diseños experimentales.

La *Estadística Descriptiva*, es una parte de la Estadística como su nombre lo indica caracteriza, detalla, especifica, grafica, presenta los datos de un conjunto de elementos delimitados espacial y temporalmente, por lo tanto, describe los elementos de una muestra objeto de investigación, resumiendo los resultados numéricamente por medio de tablas, gráficas, indicadores, estadísticos, para enunciar lo obtenido en una muestra fenómenos, temas o problemas en estudio.

La *Estadística Inferencial*, es una parte de la Estadística que hace referencia a los procedimientos y técnicas de muestreo, diseños experimentales, inferencias y predicciones de una población, para hacer pruebas de hipótesis, estimaciones, correlaciones, regresiones, modelamiento de datos.

Método Estadístico y Método Científico

El Método Estadístico son los siguientes:

1. Tema de interés o problema.
2. Planificación de la investigación.
3. Objetivos de la investigación.
4. Variables e hipótesis de investigación.
5. Población y muestra.
6. Recolección y codificación de datos.
7. Análisis e interpretación de datos.
8. Presentación de los resultados.
9. Elaboración del reporte de investigación.

Según Bertrand Russell el Método Científico puede resumirse en los siguientes 3 pasos:

1. Exacta observación del fenómeno que se estudia.
2. Formulación de una hipótesis, mediante la cual pueda explicarse los hechos observados.
3. Verificación de la hipótesis mediante nuevas observaciones.

Ejemplo 1. Para una patología que afecta a una población:

- KOCH observó ciertas formas bacilares en los esputos de pacientes tuberculosos (Paso 1).
- Como hipótesis de trabajo atribuyó a ellas la causa de la enfermedad (Paso 2).
- Comprobó que el bacilo se encontraba en los esputos de individuos tuberculosos y nunca de los procedentes sin la enfermedad (Paso 3).

Ejemplo 2. En un diagnóstico clínico:

- Se recogen los datos sobre los antecedentes del enfermo, los cuales se complementan por inspección, palpación, percusión o auscultación. Es decir, se hacen determinadas observaciones (Paso 1).
- Se hace el diagnóstico, es decir una hipótesis de trabajo (Paso 2).
- Se verifica el diagnóstico, mediante nuevas observaciones, con la ayuda del laboratorio y de otras técnicas especializadas (Paso 3).

En los ejemplos 1 y 2, se puede determinar que la Estadística interviene en el *primer paso* de la investigación científica ayudando a que las observaciones sean fidedignas y exactas. En el *segundo paso,* para formular hipótesis adecuadas, empleando un lenguaje matemático. En el *tercer paso,* para comprobar las hipótesis a través de métodos estadísticos.

Relación de la Estadística con otras Disciplinas Médicas

La Estadística es de gran utilidad médica, debido a que le permite al investigador que trata de probar una hipótesis o extraer ciertas deducciones de las observaciones realizadas:

- Decidir sobre el número de pacientes que debe estudiar para que sus conclusiones tengan validez.
- Recoger adecuadamente los datos.
- Resumir y analizar los datos recogidos.
- Evaluar más objetivamente la evidencia de otras investigaciones, comprendiendo el alcance y limitaciones.

El diagnóstico clínico individual de cualquier enfermedad, sólo es posible llegar mediante el análisis estadístico de un conjunto de signos y síntomas observados en muchos individuos. Entonces, en un pronóstico se aplica las probabilidades, por ejemplo en un paciente con Fiebre Tifoidea, predecimos que casi con seguridad se salvará, porque cuando la enfermedad se trata correctamente, sólo es fatal en aproximadamente el 3% de los casos. Por lo antes mencionado, todo nuevo tratamiento, necesita de un ensayo experimental, que se basa en la Estadística para determinar, si el mismo es realmente efectivo o inocuo.

En el campo de la *Salud Pública*, porque mediante procedimientos estadísticos se podrá estudiar la composición y las principales características de la población que se va a servir, los cambios que acontecen en ella, los riesgos a que está sometida y las necesidades que presenta. La planificación de las actividades de Salud Pública, el control de los programas que se están desarrollando y la evaluación final de sus rendimientos, eficiencia, eficacia y efectividad podrá llevarse a cabo mediante procedimientos estadísticos.

Se aplica en Demografía (ciencia que tiene como objetivo el estudio de las poblaciones humanas y que trata de su dimensión, estructura, evolución y características generales); porque estudia estadísticamente la estructura y la dinámica de las poblaciones, así como los procesos concretos que determinan la formación, la conservación y la desaparición de las poblaciones. Tales procesos, en su forma más agregada, son los de fecundidad, mortalidad y migración, emigración e inmigración.

En cuanto a la Epidemiología, como disciplina científica que estudia la distribución, la frecuencia, los determinantes, las relaciones, las predicciones y el control de los factores relacionados con la salud y con las distintas enfermedades existentes en poblaciones humanas o veterinarias específicas, lo cual se basa en procedimientos descriptivos e inferenciales para su toma de decisiones.

Referencias

Camel, F. (1991). *Estadística Médica y Planificación de la Salud.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Daniel, W. (2010). *Bioestadística: Base para el análisis de las ciencias de la salud* (4a. Ed.). México: Limusa Wiley.

Glass, G. y Stanley J. (1986). *Métodos Estadísticos Aplicados a las Ciencias Sociales.* México: Prentice Hall.

CAPÍTULO 3. CONCEPTOS BÁSICOS DE BIOESTADÍSTICA.

POBLACIÓN (N)

En términos sencillos que la población es a *quién* le vamos a indagar sobre un tema con propiedades comunes. De manera técnica, es el conjunto de elementos que cumplen ciertas propiedades comunes. Las poblaciones se definen de acuerdo al campo de interés, por lo general, se delimitan especial y temporalmente. Puede ser *finita*, cuando se tiene un número fijo de elementos o *infinita*, cuando existe un número indeterminado de elementos.

Ejemplo:

- *Población finita:* El conjunto de neonatos según el sexo en el Hospital Universitario de Los Andes, durante el último año.
- *Población infinita:* El conjunto de neonatos según el sexo en Venezuela, durante el último año.

MUESTRA (n)

Es un subconjunto de elementos de la población que cumplen ciertas propiedades comunes. En otras palabras, es parte de la población. *Ejemplo:* 100 neonatos según el sexo en el Hospital Universitario de Los Andes, durante el último año.

DATO (X_i)

Son cada uno de los elementos, individuos, cosas o entes abstractos que integran una *muestra* determinada.

UNIDAD ESTADÍSTICA

Son cada uno de los elementos, individuos, cosas o entes abstractos que integran una *población* determinada.

ESTADÍSTICO Y PARÁMETRO

Los resultados de un estudio de carácter científico se obtienen a partir de cada uno de los elementos (individuos o unidades experimentales) que son parte de la población en estudio. El valor numérico obtenido a partir de los datos individuales de los integrantes de una población, se conoce con la denominación de **parámetro**, por lo tanto este resultado muestra el comportamiento del total de datos que constituyen una población. Por lo general el objetivo de una investigación es obtener un parámetro que evalúe la situación de una población para una variable específica, por ejemplo: estado de salud, el nivel educativo, etc.

Debido a la dificultad de contar con la totalidad de los registros correspondientes a la población de interés científico, los cálculos suelen realizarse a una parte de la población, es decir una muestra y a este resultado de resumen, en terminología estadística, se le denomina un **estadístico**, el cual permite la estimación de un parámetro y la explicación del fenómeno estudiado en la población a través de un proceso de muestreo.

ESTADÍSTICO

Función definida sobre los valores numéricos de una muestra. En consecuencia, se puede expresar como el valor numérico que describe una característica o variable de la muestra y se obtiene mediante la manipulación de los datos. *Ejemplo:* Suponga que se tomó una muestra representativa de estudiantes de la Escuela de Medicina (ULA, Mérida). Para esta se calculó edad promedio y porcentaje de estudiantes que fuman, dichos resultados son estadísticos.

PARÁMETRO

Función definida sobre los valores numéricos de una población. Por lo tanto, es un valor numérico describe una característica o variable de la población. Los parámetros se obtienen a partir de la información aportada por la muestra de una población. *Ejemplo:* Si se considera la edad promedio y porcentaje de estudiantes que fuman, para el conjunto de estudiantes de la Escuela de Medicina (ULA, Mérida), se está haciendo referencia a un parámetro.

EJERCICIO 1. Determine tipo de población, población, muestra, dato, unidad estadística, estadístico y parámetro del siguiente enunciado:

Un profesor de la Escuela de Medicina de la Universidad de Los Andes (Mérida, Venezuela) durante el presente año, quiere indagar si la edad (en años cumplidos) , la clase social (alta, media, baja), el género (femenino, masculino), el nivel educativo de la madre (primaria, secundaria, superior); tienen influencia en el promedio aritmético de notas (puntos) de los estudiantes de primer año, para ello toma al azar una sección conformada por 43 estudiantes (dichos datos los recopila a través de la Oficina de Registros Estudiantiles). Se obtuvo que 60% de los 43 estudiantes son de sexo femenino y que 20 años es la edad promedio de los estudiantes de la Escuela de Medicina.

SOLUCIÓN

- **Tipo de Población:** Finita.
- **Población:** El conjunto de estudiantes de la Escuela de Medicina de la Universidad de Los Andes, Mérida, Venezuela, durante el presente año.
- **Muestra:** El conjunto de estudiantes de primer año de la Escuela de Medicina de la Universidad de Los Andes, Mérida, Venezuela, durante el presente año (n=43).
- **Dato:** Cada uno de los estudiantes de primer año de la Escuela de Medicina de la Universidad de Los Andes, Mérida, Venezuela, durante el presente año.
- **Unidad Estadística:** Cada uno de los estudiantes de la Escuela de Medicina de la Universidad de Los Andes, Mérida, Venezuela, durante el presente año.
- **Estadístico:** El 60% de los 43 estudiantes son de sexo femenino
- **Parámetro:** 20 años es la edad promedio de los estudiantes de la Escuela de Medicina.

ESCALA: Sucesión ordenada por grado o intensidad, de cosas distintas, pero de la misma especie (Real Academia Española, 2002). *Ejemplos:* Escala de colores, Escala musical, Escala de Ph.

MEDICIÓN: La asignación de categorías o números a objetos o eventos de acuerdo a un conjunto de reglas.

ESCALAS DE MEDICIÓN: Es un conjuntos de reglas que sirven para asignar categorías o números a una o varias propiedades de las cosas o eventos que se estudian. A continuación se discutirán las principales, así como las implicaciones en Bioestadística, dentro de ellas tenemos: Escala Nominal, Escala Ordinal, Escala de Intervalo, Escala de Razón.

ESCALA NOMINAL: Es una clasificación categórica no ordenada de las cosas o eventos que se estudian, por lo tanto, sólo se permite la diferencia entre categorías. Las categorías (es cada uno de los conjuntos básicos en los que puede clasificarse cada variable) son *mutuamentes excluyentes y colectivamente exhaustivas.* También están incluidas variables dicotómicas (no ordenables), ejemplo: sano o enfermo, si o no. La práctica de utilizar números para distinguir entre diversos diagnósticos médicos constituye una medición sobre una escala nominal.

ESCALA ORDINAL: Es una sucesión de tipo categórica de las cosas o eventos que se estudian. Este tipo de escala, no sólo permite la diferencia de categoría a categoría, sino que además pueden ordenarse por grados de acuerdo con algún criterio de orden (Glass y Stanley, 1986). *Ejemplos:* Niveles de una enfermedad, Rango académico, Edad (menor igual a 18 años; mayor a 18 años y menor a 40 años; mayor igual a 40 años).

ESCALA DE INTERVALO: Es una sucesión de tipo cuantitativa de las cosas o eventos que se estudian. No sólo distingue orden entre categorías, sino que también pueden discernirse diferencias iguales entre las observaciones. Se considera unidad de medida, según un parámetro (escalas de grados en temperatura, metros, pie, puntajes). Se considera un **cero arbitrario***,* es decir, el valor cero no indica ausencia de la característica, en otras palabras, la característica está presente y vale cero algunos ejemplos son la temperatura, pruebas de coeficiente intelectual, académicas, altura sobre el nivel del mar (Armas, 1988).

ESCALA DE RAZÓN: Es una sucesión de tipo cuantitativa de las cosas o eventos que se estudian. Se toma en cuenta **cero absoluto**, es decir, el valor cero representa ausencia de la característica o atributo (Armas, 1988). Claros ejemplos de esta escala son la distancia, altura, masa, peso, estatura, entre otros.

VARIABLE: Es una característica que toma diferentes valores en los elementos de la muestra o población en estudio, por lo tanto no es constante. En otras palabras, es una propiedad o característica que puede fluctuar y cuya variación es susceptible de medirse u observarse. *Ejemplos:* Presión sanguínea diastólica, Masa de niños en edad preescolar, Frecuencia cardiaca, Estatura de varones adultos, Edad de los pacientes de un médico. Los tipos de variables son: cualitativas nominales, cualitativas ordinales, cuantitativas discretas y cuantitativas continuas.

VARIABLES CUALITATIVAS: Son aquellas variables estadísticas que clasifican el conjunto de elementos de la muestra o población en categorías. *Por ejemplo:* estado civil, nacionalidad, religión, nivel educativo, grupo étnico, etc. *Las* **variables cualitativas en escala nominal**, poseen categorías no ordenadas, bien sea dicotómicas o politómicas. *Las* **variables cualitativas en escala ordinal**, consideran categorías ordenadas dicotómicas o politómicas.

VARIABLES CUANTITATIVAS: Son aquellas variables estadísticas que miden numéricamente el conjunto de observaciones de la muestra o población. *Ejemplos:* Estatura de varones adultos, Número de pupitres en un salón de clases. Las **variables cuantitativas discretas** no cuantifican valores intermedios, es decir, que entre dos valores consecutivos en la escala de la variable no existe otro valor. Las **variables cuantitativas continuas**, cuando existen valores intermedios entre dos valores consecutivos de la variable que se está midiendo.

RELACIÓN ENTRE ESCALAS Y VARIABLES

Escala Nominal - Variable Cualitativa Nominal.

Escala Ordinal - Variable Cualitativa Ordinal.

Escala de Intervalo y Escala de Razón - Variables Cuantitativas (discreta o continua).

VARIABLES DE INVESTIGACIÓN

-VARIABLE INDEPENDIENTE: Es una propiedad o característica que no depende de otra, por lo general, es el objeto o evento en el que se centra la investigación.

-VARIABLE DEPENDIENTE: Es una propiedad o característica que depende de otra o está subordinada a otra variable (variable independiente).

-**VARIABLES INTERVINIENTES:** Son variables que pueden tomar parte en el estudio o investigación.

EJERCICIO 2. Determine las variables de investigación y estadísticas de los problemas.

Problema 1. Un médico desea determinar la influencia de la Edad (años cumplidos) sobre las Enfermedades cardiovasculares, en los habitantes de Pueblo Nuevo del Sur, Mérida.

Solución del Problema 1

Variables de investigación

- *Variable independiente:* Edad (años)
- *Variable dependiente:* Enfermedades cardiovasculares
- *Variables intervinientes:* Calidad de alimentación (Buena, Regular, Mala), Ejercicio Físico (horas)

Variables estadísticas

Edad	Variable cuantitativa continúa. Escala de Razón.

Enfermedades cardiovasculares	Variable cualitativa nominal. Escala Nominal.
Calidad de alimentación	Variable cualitativa ordinal. Escala Ordinal.

Problema 2. Un investigador desea determinar si la Miopía influye en el Rendimiento promedio (puntos) de los Estudiantes de primer año de la Escuela de Medicina (Universidad de Los Andes, Mérida).

Solución del Problema 2

Variables de investigación

- *Variable independiente:* Presencia de Miopía (si, no)
- *Variable dependiente:* Rendimiento promedio (puntos).
- *Variable interviniente:* Cumplimiento del tratamiento (si, no)

Variables estadísticas

Presencia de miopía	Variable cualitativa nominal. Escala Nominal.
Rendimiento promedio	Variable cuantitativa continúa. Escala de Razón.
Cumplimiento del tratamiento	Variable cualitativa nominal. Escala Nominal.

Referencias

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

Daniel, W. (2010). *Bioestadística: Base para el análisis de las ciencias de la* salud (4a. Ed.). México: Limusa Wiley.

Glass, G. y Stanley, J. (1986). *Métodos Estadísticos Aplicados a las Ciencias Sociales.* México: Prentice Hall.

Macchi, R. (2005). *Introducción a la Estadística en Ciencias de la Salud.* Buenos Aires: Medica Panamericana.

Real Academia Española (2002). *Diccionario de la Lengua Española* (22a. Ed.). Madrid: Diccionarios Espasa.

CAPÍTULO 4. REPRESENTACIÓN Y ORGANIZACIÓN DE DATOS

Luego de estudiar los conceptos básicos podemos realizar un estudio descriptivo de una colección de datos, que habitualmente inicia con la construcción de tablas y/o gráficos adecuados al tipo de variable. La **TABLA**, consiste en presentar los datos organizadamente de manera rectangular en filas y columnas. La presentación tabular constituye la forma más general de presentar datos estadísticos (Armas, 1988). Componentes o elementos básicos de una tabla son: número, título, cuerpo y fuente.

El **número** va de acuerdo al orden de aparición en el texto.
El **título** se elabora dando respuesta al qué (variable), quién (muestra), dónde (lugar) y cuándo (mes y año).
En el **cuerpo** debe estar identificada cada una de las columnas utilizadas, iniciando con la columna de las categorías o valores que toma la variable y las columnas que siguen son los indicadores que se consideran.
La **fuente**, es de dónde se tomó la información y cuándo fue consultada.

DISTRIBUCIONES DE FRECUENCIAS UNIVARIABLE: son arreglos tabulares que consideran una variable con categorías o valores mutuamente excluyentes y colectivamente exhaustivos, muestra cómo se dividen los datos de una muestra en la variable organizada. Para la inclusión de indicadores en la tabla es menester tener claro el tipo de variable, porque de acuerdo al tipo de variable se pueden calcular los indicadores, por lo tanto, para las variables medidas en **escala nominal** solamente se le pueden hallar frecuencia absoluta, proporción, porcentaje, razón e índice (estos dos últimos no los vamos a incluir en la tabla), mientras que para las variables medidas en **escala ordinal** en adelante, se le agrega a los indicadores antes mencionados, la frecuencia acumulada, proporción acumulada y porcentaje acumulado.

Resulta oportuno mencionar algunas definiciones básicas de indicadores:

INDICADORES: Son medidas que cuantifican o cualifican dimensiones las variables en estudio.

FRECUENCIA ABSOLUTA (ni): Cantidad de elementos que existen por categoría o valor, la suma de la frecuencia absoluta debe ser igual al total de datos de la muestra.

FRECUENCIA ACUMULADA (Ni): Cantidad de elementos que existen por cada categoría o valor de manera acumulada, es decir que se va sumando de manera ordenada, la última frecuencia acumulada debe ser igual al total de la muestra.

PROPORCIÓN (P): Es una relación entre una parte con respecto al total de la muestra, también se puede establecer como la división de la frecuencia absoluta con el total de la muestra. Su valor se encuentra entre 0 y 1.

PROPORCIÓN ACUMULADA (Pa): Es una relación entre la frecuencia acumulada y el total de la muestra. Su valor se encuentra entre 0 y 1.

PORCENTAJE (%): Es una proporción multiplicada por 100, comúnmente se dice *tanto por ciento*, donde por ciento significa de cada 100 unidades.

PORCENTAJE ACUMULADO (%a): Es la proporción acumulada multiplicada por 100, indica el porcentaje que existe hasta la categoría o valor que se analiza.

RAZÓN: es la relación entre dos eventos independientes. El número obtenido indica la cantidad de veces del numerador con respecto al denominador.

ÍNDICE: es la razón multiplicada por una potencia de base 10.

Ejemplo de una tabla para una variable medida en escala nominal con todos sus elementos:

Tabla 1. Sexo de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.

Sexo	*Frecuencia absoluta (ni)*	*Proporción (P)*	*Porcentaje (%)*
Masculino	1500	0,375	37,5
Femenino	2500	0,625	62,5
Total	4000	1	

Fuente: Datos inéditos, Mayo de 2018.

Ejemplo de una tabla para una variable medida en escala ordinal con todos sus elementos:

Tabla 2. Nivel Socioeconómico de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.

Nivel Socioeconómico	*ni*	*Ni*	*P*	*Pa*	*%*	*%a*
Alto	1600	1600	0,40	0,40	40	40
Medio	1400	3000	0,35	0,75	35	75
Bajo	1000	4000	0,25	1	25	100
Total	4000	-------	1	------	100	------

Fuente: Datos supuestos, Mayo de 2018

DISTRIBUCIÓN DE FRECUENCIAS DE DATOS NO AGRUPADOS: son arreglos tabulares que consideran los datos de la variable de manera original, es decir, no requieren de una reorganización o acomodo específico. Cabe señalar que no es necesario agrupar los datos cuando existe poca variación entre los mismos.

DISTRIBUCIÓN DE FRECUENCIAS PARA DATOS AGRUPADOS: son arreglos tabulares que por lo general se utilizan para variables que presentan mucha variabilidad entre sus datos. Cuando son variables cualitativas se agrupan categorías, cuando son variables cuantitativas se agrupan en intervalos.

Cuando existen gran cantidad de datos cuantitativos (discretos y continuos) que se encuentran muy dispersos, las distribuciones de frecuencias sin agrupar no son la mejor opción para realizar una organización de datos, por lo cual se hace necesario realizar una distribución en intervalos o clases, que hagan posible un resumen de los datos de la variable en estudio, para de esta manera concentrar los datos y así acumular el número de observaciones o frecuencias contenidas para cada clase facilitando su presentación, además de permitir un análisis de aspectos resaltantes que serían muy difícil de observar con datos individuales.

Es preciso aclarar que dichas clases, deben ser mutuamente excluyentes y colectivamente exhaustivas, lo primero significa que las clases no deben estar solapadas, es decir, un valor no puede pertenecer a dos clases de manera simultánea; lo segundo expresa que todos los datos deben estar incluidos en los intervalos definidos.

Cabe señalar que, las distribuciones de frecuencias en intervalos tienen como principal desventaja, la perdida de individualidad de los datos, debido a que se sabe que en determinada clase está contenida cierta cantidad de datos, sin embargo no se conoce con exactitud los valores que toma, por lo tanto se pierde el nivel de detalle y accesibilidad. Esta desventaja es atribuible cuando los datos ya se encuentran organizados en intervalos, de lo contrario, es posible retornar a los datos originales.

Por otro lado es importante tomar en consideración que, al agrupar los datos cuantitativos en intervalos, se debe elegir un número razonable de clases, porque cuando se escoge un número muy grande el objetivo de simplificación no se obtiene, además de que se puede correr el riesgo de tener muchas clases con muy pocos datos; en el caso contrario, (se selecciona un número muy pequeño de intervalos), se resume tanto los datos al punto de perder información de utilidad. Hay que recordar que tanto el número de clases como las amplitudes de las mismas dependen de la naturaleza de los datos, el número de datos disponibles para la agrupación y el interés del investigador.

Veamos un **ejemplo** del procedimiento para agrupar datos cuantitativos en intervalos de amplitud constante, supongamos que tenemos un con conjunto de estudiantes de la Escuela de Medicina tomados al azar durante el año corriente a los cuales se les pregunta la Edad (años):

Datos:

19	25	19	26	18	32
21	27	21	25	20	29
20	26	22	24	20	24
22	22	30	22	22	24
23	23	24	28	21	31

Procedimiento para el agrupamiento de los datos:

1. **Rango Empírico (Re):**

Re= (Valor máximo – Valor mínimo) + Unidad = (32 - 18) + 1 = 14 + 1 = 15

2. **Número de clases (Nc):**

$$Nc = \sqrt{n} = \sqrt{30} = 5{,}47 \approx 5$$

3. **Amplitud del Intervalo (a_i):**

$$a_i = \frac{\text{Re}}{\text{Nc}} = \frac{15}{5} = 3$$

4. **Límite Aparente Inferior (LAI):**

   LAI = Valor mínimo = 18

5. **Límite Aparente Superior (LAS):**

   LAS = (LAI + a_i) – Unidad = (18 + 3) – 1 = 21 – 1 = 20

6. **Límite Real Inferior (LRI):**

   LRI = LAI – (Unidad/2) = 18 – (1/2) = 18 – 0,5 = 17,5

7. **Límite Real Superior (LRS):**

   LRS = LAS + (Unidad/2) = 20 + (1/2) = 20 + 0,5 = 20,5

Tabla 3. Edad (años) de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2085.

Nc	*LRI - LRS*	*LAI - LAS*	*ni*	*Ni*	*P*	*Pa*	*%*	*%a*
1	[17,5 – 20,5)	18 – 20	6	6	0,20	0,20	20	20
2	[20,5 – 23,5)	21 – 23	10	16	0,33	0,53	33	53
3	[23,5 – 26,5)	24 – 26	8	24	0,27	0,80	27	80
4	[26,5 – 29,5)	27 – 29	3	27	0,10	0,90	10	90
5	[29,5 – 32,5)	30 – 32	3	30	0,10	1	10	100
	Total		30	------	1	-------	100	-------

Fuente: Datos supuestos, Mayo de 2018.

DISTRIBUCIÓN DE FRECUENCIAS BIVARIABLE: son arreglos tabulares que consideran dos variables categóricas, bien sea ambas nominales, o una nominal y la otra ordinal, o ambas ordinales. Este tipo de organización de datos es de utilidad cuando se quieren estudiar a sus elementos según dos características, se pueden expresar con indicadores absolutos o relativos (proporción o porcentaje), cabe agregar que este tipo de tablas es de utilidad en Medicina, pues se pueden relacionar las variables, de manera inferencial por medio de un contraste de hipótesis empleando la Distribución Chi-cuadrado, además sirve para calcular indicadores de riesgo y realizar contrastes de hipótesis a través de Riesgo Relativo. Los elementos son los mismos que los señalados anteriormente, aunque es recomendado indicar si los elementos de cruce son absolutos o son relativos, además se tiene que agregar que los totales por filas y por comunas son llamados totales marginales y el total que general es el que aparece en la parte inferior derecha. Como ejemplos de tabla bivariada, de dobe entrada o contingencia se muestra la Tabla 4 y la Tabla 5:

Tabla 4. Sexo y Hábito de fumar de los estudiantes de medicina en cifras absolutas. Mérida, Venezuela. Mayo, 2018.

Sexo / Hábito de fumar	*Si*	*No*	*Total*
Masculino	1100	1000	2100
Femenino	900	500	1400
Total	2000	1500	3500

Fuente: Datos inéditos, Mayo de 2018.

Tabla 5. Nacionalidad y Ocupación de los usuarios del servicio de Odontología del IPASME en porcentajes. Mérida, Venezuela. Mayo, 2018.

Nacionalidad / Ocupación	*Docente*	*Administrativo*	*Obrero*	*Total*
Venezolano	20	30	25	75
Extranjero	5	10	10	25
Total	25	40	35	100

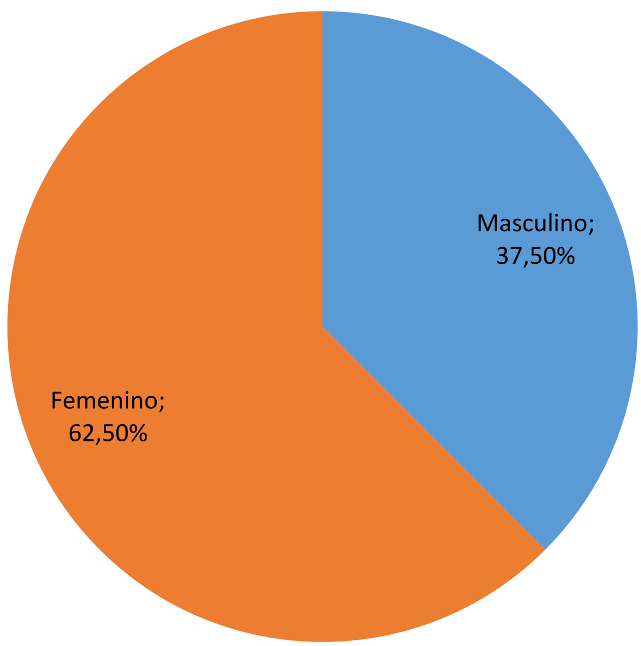
Fuente: Datos inéditos, Mayo de 2018.

GRÁFICOS: son representaciones que trasmiten en forma inmediata una idea general sobre los principales aspectos de los datos pero no proporcionan detalles. Un gráfico debe ser sencillo, de fácil interpretación y sólo debe suministrar valores relativos bien sea proporcionales o porcentuales. Son muy variados los tipos de gráficos, y por este motivo se realiza una selección de los más utilizados en Medicina (Armas, 1988). Los gráficos poseen los mismos elementos de las tablas y deben ser mostrados en dos dimensiones.

GRÁFICO CIRCULAR O DE SECTORES: se recomienda para variables cualitativas nominales y cuantitativas discretas con pocas categorías (máximo cuatro). Se construye bajo la base de dividir el círculo en sectores $\left(S_i\right)$, en tantos sectores como modalidades tenga la variable, de tal manera que cada sector sea proporcional al porcentaje datos de la correspondiente modalidad. Es necesario determinar cuántos de los trescientos sesenta grados del círculo corresponden a cada sector, para lo cual se calcula de la siguiente manera: (Armas, 1988). Se hallará como ejemplo del gráfico la variable sexo (Tabla 1).

$$Grados\ del\ Sector_{masculino} = Proporción\ x\ 100 = 0{,}375x360 = 135°$$
$$Grados\ del\ Sector_{femenino} = Proporción\ x\ 100 = 0{,}625x360 = 225°$$

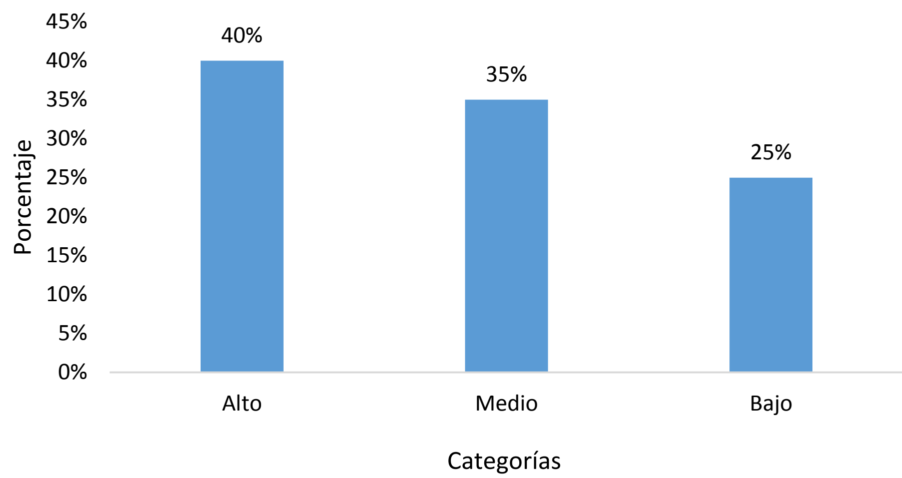
Gráfico 1. Sexo de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.



Fuente: Datos Supuestos, Mayo de 2018.

GRÁFICO DE BARRAS SIMPLES: se recomienda para datos medidos en escala nominal y ordinal. Para construir este gráfico, se utiliza en el eje de las abscisas o eje "X" las categorías y en el eje de las ordenadas o eje "Y", la proporción o porcentaje de la variable o frecuencia absoluta de cada categoría. Para cada categoría se levanta una barra en forma rectangular cuya altura viene dada por el número de datos de cada categoría. Las barras deben ir separadas y tanto el ancho como la distancia que las separa son arbitrarios, pero una vez fijados deben mantenerse en todo el gráfico (Armas, 1988). Se hallará como ejemplo del gráfico la variable nivel socioeconómico (Tabla 2).

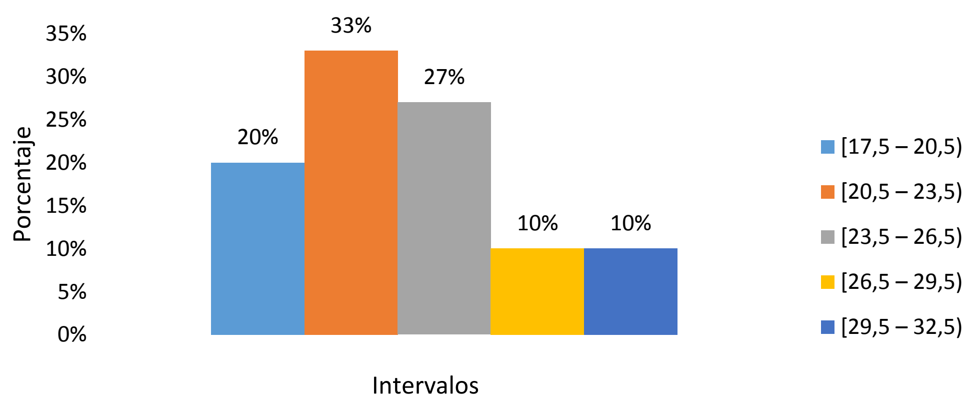
Gráfico 2. Nivel Socioeconómico de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.



Fuente: Datos supuestos, Mayo de 2018

HISTOGRAMA: se utiliza para representar distribuciones de frecuencias cuyas clases son intervalos. Consiste en un gráfico de barras rectangulares con la particularidad de que las barras están juntas unas de otras y el ancho de la misma es arbitrario, pero una vez fijada deben mantenerse en todo el gráfico. Se construye llevando sobre el eje de abscisas los límites reales de las clases y en el eje de las ordenadas cifras relativas (proporción o porcentaje). Se hallará como ejemplo del gráfico la variable Edad (Tabla 3).

Gráfico 3. Edad (años) de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.



Fuente: Datos Supuestos, Mayo de 2018.

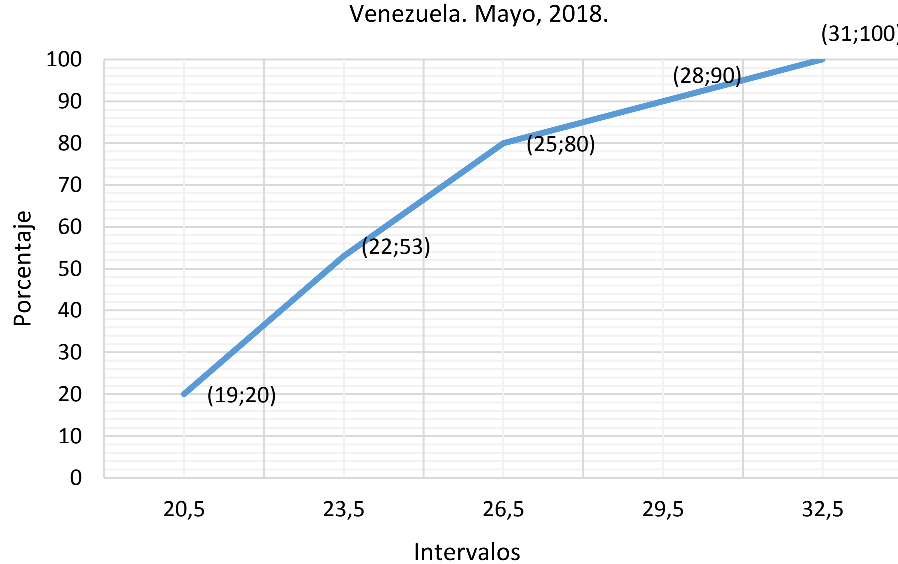
POLIGONOS DE FRECUENCIAS: Es un tipo de curva que se utiliza para representar distribuciones de frecuencias cuyas clases son intervalos, en su construcción intervienen las marcas de clase (Xm) o puntos medios (se obtienen de sumar los límites aparentes y dividirlos entre dos), las cuales se llevan en el eje de las abscisas y las proporciones o porcentajes se colocan en el eje de las ordenadas.

Se intercepta cada marca de clase con su correspondiente punto medio, es decir, se genera un par ordenado de coordenadas (Xmi, Pi) o (Xmi, %i), posteriormente se unen los pares ordenados en forma ordenada por medio de segmentos de recta. Para cerrar esa línea poligonal con el eje de abscisas, se crea una clase ficticia anterior a la primera clase con su misma amplitud y se prolonga la poligonal hasta el punto medio de esa nueva clase; se procede de manera similar con la última clase y se obtiene la curva llamada polígono de frecuencias.

Un procedimiento alternativo para obtener el polígono de frecuencias, es a partir del histograma, utilizando los puntos medios de los rectángulos y para cerrar la curva se procede de la misma manera como se indicó en el parágrafo anterior.

Ejemplo a partir de la información de la Tabla 3

Gráfico 4. Edad (años) de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.

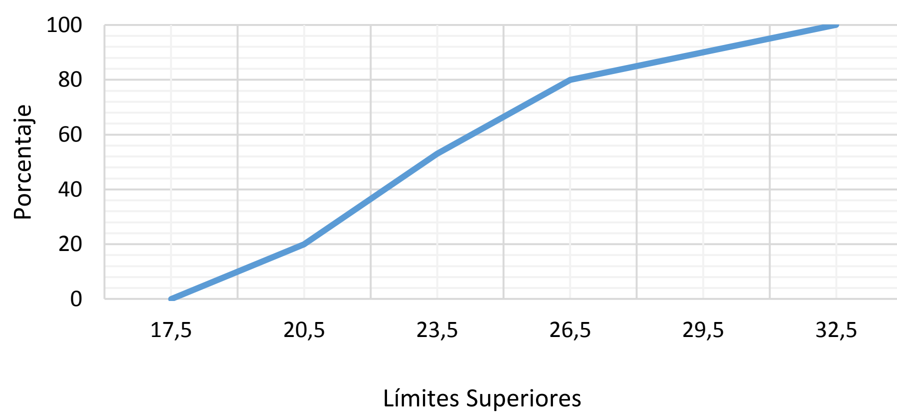


Fuente: Datos supuestos, Mayo de 2018.

OJIVA O POLÍGONO DE FRECUENCIAS ACUMULADAS: Es un gráfico curvilíneo para representar las frecuencias acumuladas de una distribución de datos cuyas clases son intervalos. Se obtiene llevando sobre el eje horizontal de un plano cartesiano lo límites superiores de cada clase y sobre el eje vertical las proporciones o porcentajes acumulados de la clase. Se intercepta cada límite superior con su correspondiente proporción o porcentaje acumulado, además se une el primer límite inferior con la proporción o porcentaje acumulado igual a cero. Se unen todos los puntos con segmentos de recta y se obtiene una curva creciente denominada ojiva porcentual o polígono de frecuencias acumuladas.

Ejemplo a partir de la información de la Tabla 3

Gráfico 5. Edad (años) de los estudiantes de medicina. Mérida, Venezuela. Mayo, 2018.



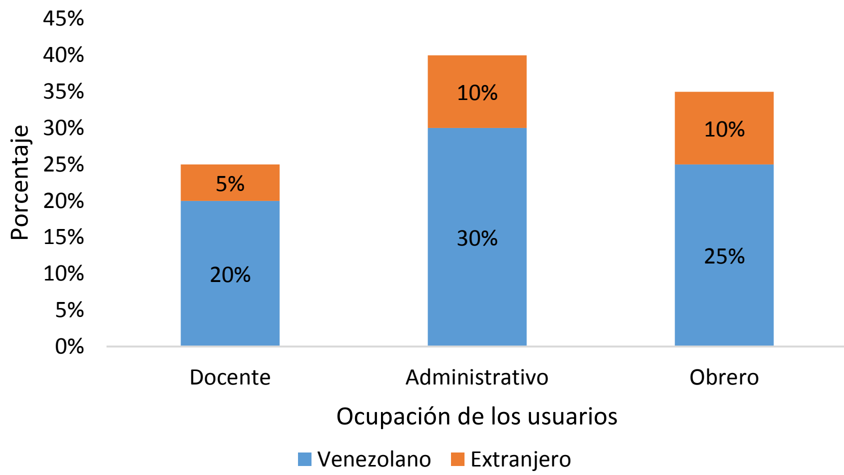
Fuente: Datos supuestos, Mayo de 2018.

Nota: solo se toma el primer LRI, que en este caso es 17,5 para intersectarlo con cero

GRÁFICO DE BARRAS APILADAS: Es un gráfico que apila una serie de rectángulos (uno sobre otro), en donde el área de cada rectángulo es proporcional a la cantidad de la muestra que representa, de tal manera que la altura de la barra resultante representa la suma de los valores de la variable. Este tipo de gráfico puede utilizarse para comparar fácilmente las partes de un todo, por lo tanto, se recomienda cuando la suma de las proporciones o porcentajes de las variables es de interés. En el eje X se coloca las categorías de una de las variables y sobre la barra las categorías de la otra variable, en el eje Y se muestra la proporción, porcentaje u otro indicador relativo que se esté manejando.

Ejemplo a partir de la información de la Tabla 5

Gráfico 6. Nacionalidad y Ocupación de los usuarios del servicio de Odontología del IPASME en porcentajes. Mérida, Venezuela. Mayo, 2018.

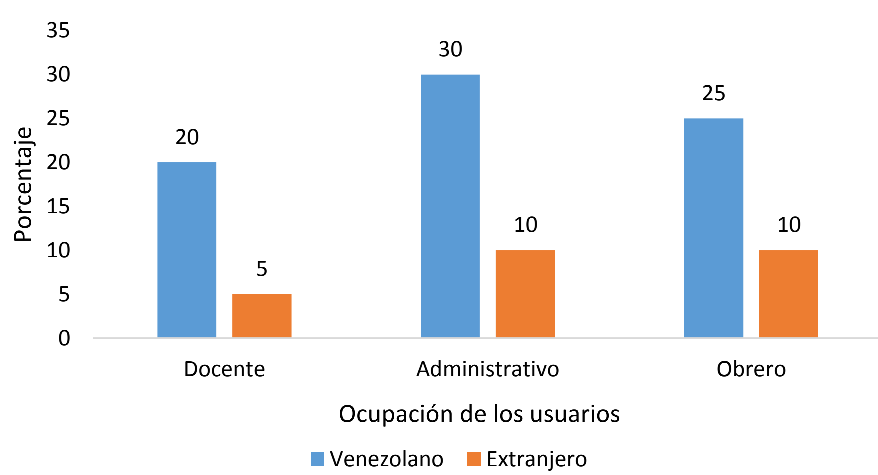


Fuente: Datos supuestos, Mayo de 2018.

GRÁFICO DE BARRAS AGRUPADAS: Es un gráfico que agrupa una serie de rectángulos (uno al lado del otro), se construye colocando en el eje de las abscisas las categorías de una de las variables y en cada categoría se muestran las barras con las categorías de la otra variable, en el eje de las ordenadas se representa la proporción, porcentaje u otro indicador relativo que se esté utilizando.

Ejemplo a partir de la información de la Tabla 5

Gráfico 7. Nacionalidad y Ocupación de los usuarios del servicio de Odontología del IPASME en porcentajes. Mérida, Venezuela. Mayo, 2018.

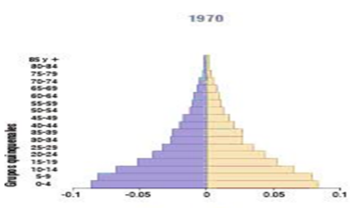


Fuente: Datos supuestos, Mayo de 2018.

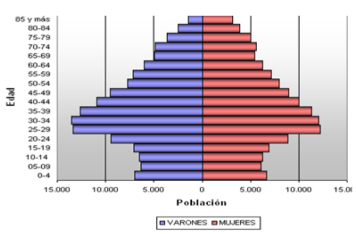
PIRÁMIDE DE POBLACIÓN: La representación gráfica de la composición de una población por edad y sexo se conoce con el nombre de pirámide de población. La pirámide poblacional es una representación en forma de histograma de la distribución por edad y sexo de una población. En este tipo de gráfico, el eje vertical representa la edad, situándose al lado izquierdo el sexo masculino y al lado derecho el sexo femenino. En el eje horizontal se refleja, en forma absoluta o relativa, la magnitud de la distribución por edad. Cada barra representa un grupo etario, por lo general se usa una amplitud de 5 años (o una edad en el caso de edades simples). La longitud de cada barra es proporcional al tamaño de cada grupo de edad por sexo, o al porcentaje que representan las personas en la población total.

La pirámide de población puede ser representada en números absolutos o en porcentajes, si se construye en base a porcentajes, el total de varones y mujeres debe usarse para calcular los porcentajes para cada edad. No debe usar la población total para calcular porcentajes para los varones por grupo de edad.

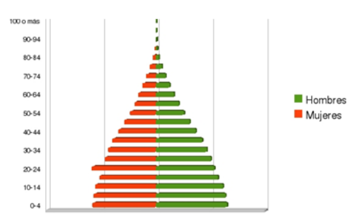
Interpretaciones de la pirámide poblacional de acuerdo a su forma:



1. Población expansiva. Mayor número de personas en las edades menores.



2. Población constrictiva. Menor número de personas en las edades menores.



3. Población estacionaria. Aproximadamente igual número de personas en todos los grupos de edades, con una reducción paulatina en las edades más avanzadas

El factor que influye en mayor grado sobre la conformación de la pirámide de población es la tasa de fertilidad. A mayor cantidad de hijos por progenitor la base de la pirámide será más ancha.

La influencia de la tasa de mortalidad no es tan simple. Al contrario de lo que se podría suponer, las poblaciones con bajas tasas de mortalidad tienen una distribución por edades algo más joven porque las diferencias en las tasas de mortalidad frecuentemente surgen de los grupos de menor edad (en especial lactantes y niños).

Las pirámides de población pueden usarse para:

• Observar el impacto bruto de la fecundidad, migración y mortalidad en la población de una localidad.

• Observar cambios en la estructura por edad y sexo cuanto las pirámides se desarrollan por lo menos en dos puntos del tiempo.

• Ayudar a los planificadores a observar cambios futuros en la población. En este caso, los planificadores pueden comparar una pirámide de población que se basa en un censo reciente con una que se basa en proyecciones de población.

GRÁFICO DE CAJA Y BIGOTE: Es una representación que describe simultáneamente varias características importantes de un conjunto de datos, tales como el centro, la dispersión, la desviación de la simetría y la identificación de observaciones que caen inusualmente lejos de la mayoría de datos (a estas observaciones se le llaman valores atípicos).

En este tipo de gráfico se representan los tres cuartiles y los valores extremos (máximo y mínimo) en una caja rectangular alineada en sentido horizontal o vertical, abarcando el rango intercuartílico, con el borde izquierdo (o el inferior) de la caja el primer cuartil (**Q1**=P25) y en el borde derecho (o el superior) el tercer cuartil (**Q3**=P75). Se traza una línea a través de la caja en el segundo cuartil (**Q2**=D5=P50=Mediana). Posteriormente se traza una línea o **bigote**, desde cada extremo de la caja.

El bigote inferior o izquierdo es una línea que va desde el primer cuartil al punto correspondiente al menor de los datos dentro de 1,5 rangos intercuartílicos a partir del primer cuartil. El bigote superior derecho es una línea que va desde el tercer cuartil al punto correspondiente al mayor de los datos dentro de 1,5 rangos intercuartílicos a partir del tercer cuartil. Los datos que se encuentran alejados de la caja más allá de los bigotes se grafican como puntos individuales. A un punto situado después de un bigote, pero a menos de 3 rangos intercuartílicos de los bordes de la caja se le llama **punto atípico**, mientras que a un punto que se encuentre situado a más de 3 rangos intercuartílicos de los bordes de la caja se le denomina **punto atípico extremo.**

Ejemplo:

Daniel (2010) menciona que en la Revista American Journal of Clinical Pathology, Pitts y otros, asegura que "*los carcinomas con metaplasia y sarcomas producidos dentro del seno son difíciles de diagnosticar y clasificar con precisión debido a sus variados patrones histológicos y a su rareza*". En un intento por estudiar más detalles de las características biológicas, los autores investigaron una serie de sarcomas puros y carcinomas que exhibían metaplasia, a continuación se muestran los datos para la construcción del gráfico de caja.

Diámetros (cm) de sarcomas puros extirpados de los senos de 20 mujeres.

0,5	1,2	2,1	2,5	2,5	3,0	3,8	4,0	4,2	4,5
5,0	5,0	5,0	5,0	6,0	6,5	7,0	8,0	9,5	13,0

Construcción del gráfico:

Se calculan los cuartiles en el programa SPSS para Windows versión 20, se hace clic en el menú *Analizar*, se selecciona *Estadísticos descriptivos* y luego *Frecuencias;* en el cuadro diálogo que arroja se selecciona la variable en estudio, se hace clic en el botón *Estadísticos...* (Ubicado en la parte superior derecha), se selecciona *Cuartiles* (ubicado en la parte superior izquierda), se hace clic en el botón *Continuar* y finalmente en *Aceptar.*

Los resultados obtenidos fueron los siguientes:

Q1=2,625 cm

Q2=4,75 cm

Q3=6,5 cm

Se calcula la Rango intercuartilar (RQ): RQ = Q3 - Q1= 6,5 cm - 2,625 cm = 3,875 cm

El bigote superior va a quedar determinado por el mayor de los datos comprendidos entre Q3 y Q3 + 1,5.RQ. Entonces se tiene que:

Q3 + 1,5.RQ = 6,5 cm + 1,5 (3,875 cm) = 6,5 cm + 5,8125 cm = 12,3125 cm
Por lo tanto el bigote superior llegará hasta el mayor de los datos entre 6,5 cm y 12,3125 cm, el cual es 9,5 cm.

El bigote inferior va a quedar determinado por el menor de los datos comprendidos entre Q1 y Q1 - 1,5.RQ. Entonces se tiene que:

Q1 - 1,5.RQ = 2,625 cm - 1,5 (3,875 cm) = 2,625 cm - 5,8125 cm = -3,1875 cm
Por lo tanto el bigote inferior llegará hasta el menor de los datos entre 2,625 cm y -3,1875 cm, el cual es 0,5 cm.

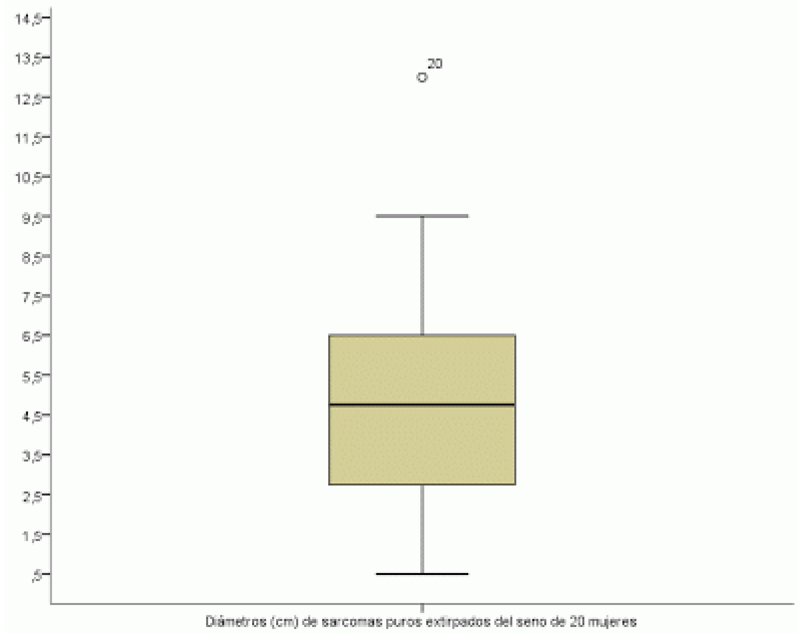
También se tiene que el sarcoma de 13 cm de diámetro es un valor o punto atípico.

Para hallar esta gráfica en el paquete estadístico SPSS para Windows versión 19, se realiza el siguiente procedimiento:

Se hace clic en el menú *Gráficos*, se selecciona *Cuadro de diálogos antiguos* y luego *Diagramas de caja...*; en el cuadro diálogo que arroja se selecciona el *Diagrama simple* y *Resúmenes para distintas variables*, se hace clic en el botón *Definir*, se coloca la variable en estudio en el recuadro *Las cajas representan* y finalmente se hace clic en el botón *Aceptar*.

El resultado es el siguiente luego de la edición de escala, se muestra en el Gráfico 8.

Gráfico 8. Diámetros (cm) de sarcomas puros extirpados de los senos de 20 mujeres.



Fuente: Daniel (2010).

Al examinar la gráfica se puede ver que el 50% central se encuentra entre 2,6 y 6,5 cm, además se puede observar que la Mediana está cercana a 4,5 cm. La línea o bigote más alargada en la parte superior indica que la distribución de diámetros está inclinada hacia la derecha o es asimétrica positiva. Asimismo, se puede notar que el programa estadístico indica cual es el dato extremo con un círculo y dice cuál es el número del mismo en la base de datos.

La *gráfica de caja* con valores extremos, así como el *diagrama de tallos y hojas*, son ejemplos de lo que se conoce como *análisis exploratorio de datos*. Estas técnicas, se hicieron populares debido al resultado del trabajo de Tuckey, permiten que el investigador examine datos de manera que éstos revelen tendencias y relaciones, identifique características únicas del conjunto de datos y faciliten su descripción y resumen.

DIAGRAMA DE TALLOS Y HOJAS: es de utilidad para representar un conjunto de datos cuantitativos, este tipo de representación presenta similitudes con el histograma en cuanto que proporciona información del recorrido de la distribución de datos en estudio, muestra la ubicación de la mayor concentración de mediciones y revela la presencia o ausencia de simetría.

Cabe señalar que el diagrama de tallo y hojas tiene ventajas sobre el histograma, porque conserva la información que puede arrojar las mediciones individuales, situación que se pierde en los intervalos del histograma, otra ventaja, es que ésta representación elimina el paso de los datos originales a clases, lo que hace notar que se puede construir en el proceso de marcaje de los elementos en estudio.

¿Cómo construir el diagrama de tallo y hojas?

- Se debe dividir cada medición en dos partes, la primera se llama *tallo* y la segunda *hojas*.
- El tallo se forma con uno o más dígitos iniciales de la medición, y las hojas se forman con uno o más de los dígitos restantes.
- La cantidad de tallos preferiblemente deben ser mayores o iguales a 5 y menores o iguales a 20.
- Los tallos forman una columna ordenada de menor a mayor del lado izquierdo del diagrama.
- Registrar las hojas por cada observación junto al valor correspondiente del tallo.

Ejemplo del diagrama de tallo y hojas:

Supongamos que durante el presente mes del año corriente, se mide el Tiempo en minutos que tardan un grupo de estudiantes para llegar a la Facultad de Medicina, Universidad de Los Andes (ULA).

Datos de Tiempo (minutos): 12, 13, 13, 14, 14, 15, 15, 20, 21, 22, 22, 22, 25, 27, 28, 29, 31, 31, 33, 34, 36, 38, 39, 40, 41, 41, 42, 44, 45, 50, 51 52.

Gráfico 9. Estudiantes de la Facultad de Medicina según Tiempo en minutos para llegar a la Facultad de Medicina. ULA, Mérida. Marzo, 2015.

Tallo	*Hoja*	*Frecuencia absoluta*
1	2,3,3,4,4,5,5	7
2	0,1,2,2,2,5,7,8,9	9
3	1,1,3,4,6,8,9	7
4	0,1,1,2,4,5	6
5	0,1,2	3

Fuente: Datos supuestos, Marzo de 2015.

Interpretación: a través del Diagrama 1 se puede observar que la distribución de datos de Tiempo (minutos) es asimétrica positiva, con la mayor concentración de datos en el segundo tallo, también se puede determinar la **moda,** que en este caso es 22 minutos. Además se puede hallar la **mediana,** considerando que la cantidad de datos son pares, se tiene que los elementos centrales son 29 y 31 minutos, por lo tanto la mediana = (29+31)/2=30 minutos, es el Tiempo que divide la distribución en dos partes iguales. Asimismo se pueden calcular los **cuartiles**, usando Q1=n/4 y para el Q3=3n/4, por ejemplo: Q1=32/4=8, el valor obtenido determina la posición, la cual arroja que 20 minutos es el Tiempo que deja el 25% de los datos por debajo y el 75% de los datos de la distribución por encima.

Recomendaciones:
Es importante tomar en cuenta que este tipo de diagramas, no es aconsejable en informes anuales o en algún tipo de medios de difusión para un público en general, porque se convierten en una ayuda básica para que investigadores y tomadores de decisiones comprendan la naturaleza de los datos.

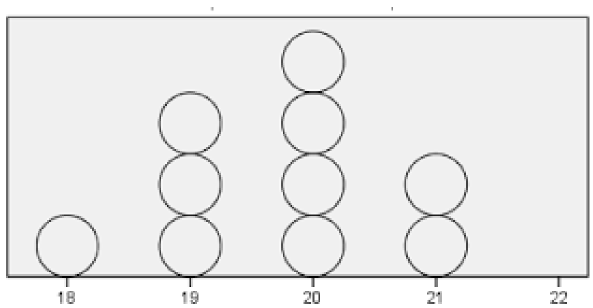
Algunas veces, la utilización del primero o de los dos primeros dígitos de los datos puntuales como tallos no proporcionan suficientes tallos como para permitir detectar la forma de su distribución. Una manera de solucionar esto es utilizar tallos dobles. Es decir, utilizar cada tallo dos veces: una vez para trazar las hojas inferiores y otra vez para trazar las hojas superiores.

Muchos de los procedimientos estadísticos que se desarrollan suponen que la variable independiente estudiada posea al menos una distribución aproximadamente normal, para la cual el diagrama de tallos y hojas tiene forma de campana, por lo tanto, los diagramas de tallos y hojas dan una idea de la localización de los datos y de la forma de la distribución. Esta técnica funciona bien para los conjuntos de datos que no tienen una dispersión muy grande.

GRÁFICO DE PUNTOS Y PICTOGRAMAS: Con la creciente potencia y disponibilidad de herramientas computacionales, cada vez se hace más fácil encontrar múltiples representaciones gráficas para describir datos. El problema se centra principalmente en seleccionar las más apropiadas para cada ocasión.

El **gráfico de puntos**, se utiliza para un conjunto pequeño de datos cualitativos y cuantitativos, este tipo de gráfico se representa en un línea recta los elementos mediante puntos. **Ejemplo de** Chipia (2012):

Gráfico 9. Edad (años cumplidos). Mérida, Venezuela. Mayo, 2018.



Fuente: Datos supuestos, Mayo de 2018.

Pictogramas: tienen su uso principal en la representación gráfica de series cronológicas y en algunos casos de datos cualitativos. Es un gráfico muy expresivo y atrapa rápidamente la atención del lector, cada símbolo utilizado representa la naturaleza de los datos. Cada símbolo equivale a un número determinado de unidades y se repite las veces que sea necesario hasta que refleje la magnitud del valor de la variable. Los pictogramas se emplean para hacer comparaciones entre valores de la variable analizada y no se utiliza para hacer afirmaciones aisladas.

Ejemplo:

Gráfico 10. Habitantes por año.

Fuente: Datos supuestos.

Referencia

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

Chipia, J. y Lara, C. (2008). *Módulo para la enseñanza–aprendizaje de la estadística en el bachillerato mediante la solución de situaciones problemas.* Mérida: Tesis de Grado mención Publicación en la Universidad de Los Andes.

Chipia, J. (2009). *Propuesta para la enseñanza de la estadística en primer año de secundaria mediante resolución de problemas.* Revista Voces: Tecnología y pensamiento, 4(1-2), 79-96.

Chipia, J. (2012). *Gráficas de caja* [blog]. Bioestadística y Matemática Básica #ULA. Disponible: http://bioestadisticaula.blogspot.com/2012/09/graficas-de-caja.html [Consulta: 2015 Marzo 30].

Chipia, J. (2012). *Gráfico de puntos y pictogramas* [blog]. Bioestadística y Matemática Básica #ULA. Disponible: http://bioestadisticaula.blogspot.com/2012/07/grafico-de-puntos-y-pictogramas.html [Consulta: 2015 Marzo 30].

Chipia, J., Cadenas, R. y Lara, C. (2012) *Propuesta para la enseñanza de organización de datos para variables cualitativas.* Revista EDUCERE, 16 (53), 185-196.

Daniel, W. (2010). *Bioestadística. Base para el análisis de las ciencias de la salud* (4a. Ed.). México: Limusa Wiley.

Montgomery, D. y Runger, G. (2008). *Probabilidad y estadística aplicada a la ingeniería* (2a. Ed.). México: Limusa Wiley.

CAPÍTULO 5. MEDIDAS DE TENDENCIA CENTRAL

Cuando se desea información sobre los valores medios de la serie de datos de la muestra, se utilizan una serie de Medidas descriptivas que tratan de representar o resumir los datos de la muestra, estas son llamadas **MEDIDAS DE TENDENCIA CENTRAL (MTC)**: las cuales indican la posición hacia la que tienden a concentrarse las observaciones o alrededor del cual se distribuyen el conjunto de datos (Glass y Stanley, 1986). A continuación se estudiaran las principales, las cuales son de uso común para describir los datos, estas Medidas son: la Moda, Mediana y Media Aritmética.

¿QUÉ ESTÁ DE MODA? Facebook, Internet, celulares, computadoras, televisión, la política, jugar fútbol, entre otros. Todo lo antes enunciado es lo que está de moda y se sabe **¿por qué está de MODA?** La mayoría de las personas quieren comprar un celular, una computadora, usar Facebook, de igual forma la mayoría de las personas quieren navegar en Internet un par de horas, o distraerse jugando fútbol un rato, o en su defecto hablar un poco de política.

A partir de lo anterior la **MODA** (**Mo**) es lo que se realiza en su mayoría, en otras palabras, constituye el nivel de rendimiento de mayor frecuencia observado en una determinada distribución o conjunto de datos, dentro de la escala correspondiente, y con un determinado margen de error por exceso o por defecto (Martínez, 2008).

En una distribución de datos puede existir una sola **MODA** (distribución unimodal), también puede tener dos **MODAS** (distribución bimodal) o más de dos **MODAS** (distribución multimodal).

Para calcular la **MODA**, se determina el valor o valores de la distribución que obtienen la frecuencia máxima.

Mo= *Valor de* X_i *con Frecuencia Máxima (uno o más)*

Ejemplo. *Datos de Edad (años cumplidos) de 6 estudiantes universitarios*: 20 años, 21 años, 21 años, 22 años, 29 años, 28 años.

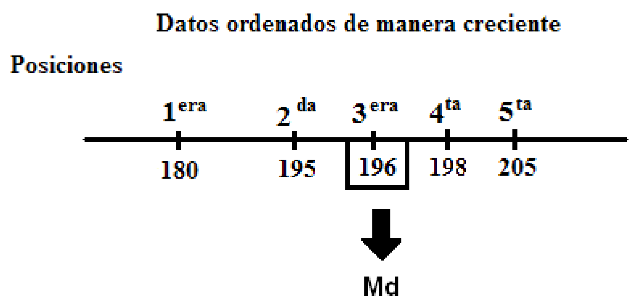
Moda (Mo) = 21 años.

Interpretación: 21 años es la Edad de mayor frecuencia de los estudiantes universitarios objeto de estudio.

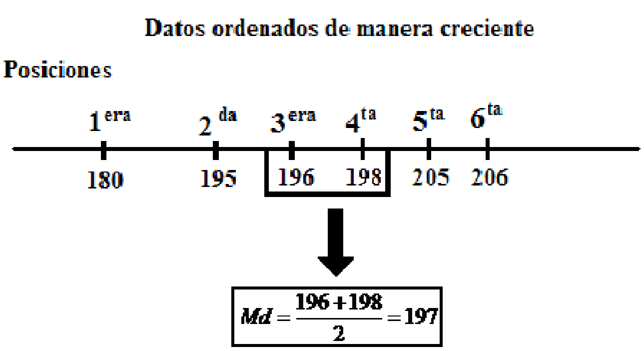
¿PARA QUÉ HALLAR LA PUNTO CENTRAL DE UN CONJUNTO DE DATOS? Para determinar el punto central, implica la obtención de un valor o categoría que deja aproximadamente el 50% de los datos por debajo y el 50% por encima.

¿CÓMO HALLAR EL PUNTO CENTRAL DE UN CONJUNTO DE DATOS? El estadístico que permite hallar el punto central de un conjunto de datos es la **Mediana (Md)**, para obtenerla se ordenan los datos en forma **ascendente o descendente** según la magnitud correspondiente, y se determina cual de esos puntajes u observaciones se encuentra en el centro de la serie ordenada de datos.

Cuando el número de observaciones o datos es **impar,** siempre existe un valor que ocupa el punto central, el cual va a ser la **Mediana (Md).** Por ejemplo:



Cuando el número de observaciones o datos es **par,** se suman los *dos valores centrales* y se divide entre dos, el resultado va a ocupar el punto central, el cual va a ser la Mediana. Por ejemplo:



La **Mediana** (Md), con datos previamente ordenados de manera ascendente es un valor que a lo sumo es menor a la mitad de los datos y a su vez es mayor a lo sumo, que la mitad de los datos. Alternativamente podemos decir, la **Mediana** ocupa la posición central de los datos, una vez ordenados. En forma aproximada podemos decir que el 50% de los datos están por debajo de la **Mediana** y el 50% por encima.

Es importante señalar que, la Mediana (Md) de un Grupo de datos es aquel punto en una línea de números, tal que la suma de las distancias absolutas de todas las puntuaciones en el Grupo hacia ese punto es menor, a la suma de la distancia de cualquier otro punto (Glass y Stanley, 1986: 68).

¿CÓMO DETERMINAR EL RENDIMIENTO CARACTERÍSTICO DE UNA SERIE DE DATOS? Para dar respuesta se debe hallar el estadístico **Media Aritmética** (**Mx**), la cual determina un valor que indica el **RENDIMIENTO CARACTERÍSTICO** (mayor o menor), correspondiente a una distribución de frecuencias o serie de datos, dentro de la escala correspondiente, y con un determinado margen de error por exceso o por defecto (Hernández, 2011). Es importante destacar que sólo puede ser calculada para variables cuantitativas.

¿CÓMO SE CALCULA LA MEDIA ARITMÉTICA? De una colección de datos $x_1, \ldots, x_N$ se define como la suma de esos datos, dividida entre el número de sumandos (N).

$$Mx = \frac{x_1 + \ldots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Ejemplo. *Datos de Edad (años cumplidos) de 6 estudiantes universitarios*: 20 años, 21 años, 21 años, 22 años, 29 años, 28 años.

$$Mx = \frac{20+21+21+22+29+25}{6} = \frac{138}{6} = 23 años$$

Interpretación: 23 años es la Edad característica de los estudiantes universitarios objeto de estudio.

Cabe agregar que la Media Aritmética en un Grupo de puntuaciones, es aquel punto en una línea de números, tal que la suma de los cuadrados de las distancias de todas las puntuaciones a ese punto, es menor a la suma de los cuadrados de las distancias a cualquier otro punto (Glass y Stanley, 1986).

Resulta oportuno mencionar que cuando los valores de **Mx=Md=Mo** la *distribución es simétrica.* Sin embargo, la mayoría de distribuciones de datos son no normales, en ese caso se puede determinar que es asimétrica positiva cuando **Mx>Md>Mo,** o es asimétrica negativa cuando **Mx<Md<Mo**.

Referencias

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

Chipia, J. y Lara, C. (2008). *Módulo para la enseñanza–aprendizaje de la estadística en el bachillerato mediante la solución de situaciones problemas.* Mérida: Tesis de Grado mención Publicación en la Universidad de Los Andes.

Chipia, J. (2014). *Efectividad de un programa de enseñanza/aprendizaje sobre estadística descriptiva.* Mérida: Tesis de Maestría mención Honorífica en la Universidad de Los Andes.

Glass, G. y Stanley, J. (1986). *Métodos Estadísticos Aplicados a las Ciencias Sociales.* México: Prentice Hall.

Hernández, R. (2011). *Variabilidad absoluta y relativa en distribuciones de frecuencias.* Mérida: Consejo de Estudios de Postgrado de la Universidad de Los Andes.

Martínez, C. (2008). *Estadística y muestreo* (12da. Ed.). Bogotá: ECOE Ediciones.

CAPÍTULO 6. MEDIDAS DE VARIABILIDAD

La Variabilidad determina cómo están dispersos los datos en estudio con respecto a un valor medio. Si todos los datos son iguales, **NO** existe variabilidad o la Variabilidad es igual a cero. Las **MEDIDAS DE VARIABILIDAD ABSOLUTA** (MVA) también llamadas de Medidas de Dispersión o Heterogeneidad Absolutas, permiten cuantificar o expresar la variación, dispersión o el grado de distanciamiento de un grupo de valores respecto a un valor medio de la serie de datos, en otras palabras representan la concentración de los datos en una determina escala numérica en un grupo (Glass y Stanley, 1986; p.75).

¿Cómo se define el Rango Empírico? El **RANGO EMPÍRICO** o simplemente **Rango** (**R**), se define como el intervalo que contiene a los datos objeto de estudio, es decir, es la fluctuación entre los valores extremos (el mayor y el menor valor) de los datos objeto de estudio (Martínez, 2008; p. 215). Este estadístico se calcula hallando la diferencia entre el **Valor Máximo** y el **Valor Mínimo** de los datos de la variable objeto de estudio así:

$$R = (ValorMáximo - ValorMínimo)$$

¿Qué inconvenientes presenta el Rango? Aunque los datos de dos distribuciones posean el mismo valor del Rango **NO** muestra cuál distribución es más representativa, confiable o estable; esta medida solo determina cuál es la diferencia entre el mayor y el menor valor de los datos objeto de estudio, por lo tanto es una medida no estable y difícil de interpretar porque depende de la escala de valores que se considera y a quienes se está estudiando, lo último coloca en relieve lo poco confiable que es esta Medida de Variabilidad Absoluta.

Ejemplo. *Datos de Edades (años cumplidos) de 6 Estudiantes Universitarios*: 20 años, 21 años, 21 años, 22 años, 29 años, 28 años.

$$R = (ValorMáximo - ValorMínimo) = 29 - 20 = 9 años$$

Lo que se ha determinado es que, 9 años es el recorrido desde el menor valor hasta el mayor valor de los datos de la variable Edad de los Estudiantes Universitarios objeto de estudio. Por lo tanto, existe un recorrido Medio entre el mayor y el menor valor de los datos.

Para determinar si este **Rango** (**R**) es *Alto, Medio o Bajo*, se debe considerar la escala de valores que puede tomar los datos de la variable que se investiga, **por ejemplo**: la Edad (años cumplidos) puede tomar valores desde 0 años hasta 115 años, entonces me pregunto **¿cuál puede ser un Rango Alto, Medio o Bajo en esta Variable**? Determinar un Nivel en el Rango es **RELATIVO,** se debe tomar en cuenta cómo y quién lo observe y en qué contexto se realiza la observación. Es decir, 9 años es un recorrido o Rango de Edad *Medio* en Estudiantes Universitarios regulares de la Universidad de Los Andes, sin embargo pudiera ocurrir que en Estudiantes de la Universidad Bolivariana de Venezuela es un Rango de Edad *Bajo* y en Estudiantes de Bachillerato regulares es un recorrido *Muy Alto*.

VARIANZA (S^2x)*:* es el promedio de las desviaciones cuadráticas de los valores de la variable respecto a la media aritmética (Armas, 1988).

Cuando n ≤ 30, se calcula la varianza sesgada, es decir dividiendo entre (n - 1), donde n - 1 representa los **grados de libertad** (el número de variaciones libres que tiene una variable aleatoria dentro de su rango):

Para datos no agrupados

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

Para datos agrupados:

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

Cuando n > 30, se calcula la varianza insesgada, es decir dividiendo entre n:

Para datos sin agrupar:

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

Para datos agrupados:

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

Un ejemplo sencillo ilustra el concepto de grados de libertad: En una bolsa de papel tenemos cinco bolas, cada una marcada con el número correspondiente de 1 a 5. Si sacamos al azar una primera bola, puede variar entre 1 y 5, si la primera bola es la 3, al sacar la segunda la bola puede variar entre 1, 2, 4 y 5. Si la bola sacada es la 5, al sacar la tercera bola, puede variar entre 1, 3 y 2. Al sacar la tercera bola y es la 1, la cuarta bola puede variar entre 3 y 2. Si la cuarta bola es la 2, la quinta bola necesariamente es la 3 (ya no puede variar). En total sólo hay n – 1 variaciones libres = 5 – 1 = 4.

Ejemplo. *Datos de Edades (años cumplidos) de 6 Estudiantes Universitarios*: 20 años, 21 años, 21 años, 22 años, 29 años, 28 años.

Calculo de la Varianza (**S^2x**): recordando que la Media Aritmética (**Mx)** es igual a **23 años**.

$$S^2x = \frac{(20-23)^2 + (21-23)^2 + (21-23)^2 + (22-23)^2 + (29-23)^2 + (28-23)^2}{6} = 13{,}1\widehat{6} \quad años^2$$

Interpretación: 13,17 años^2 es la distancia promedio en unidades cuadráticas entre cada Edad y la Mx de los Estudiantes Universitarios.

Cálculo de la Desviación Típica (**Sx**)

$$Sx = \sqrt{13{,}17} = 3{,}627671\ldots \approx 3{,}63\,años$$

La Desviación Típica, halla el valor de la raíz cuadrada positiva de la varianza, por lo cual la unidad de medida de la variable objeto de estudio se expresa en unidades lineales, lo que muestra que entre S^2x y Sx existe una diferencia de escala.

Interpretación: el distanciamiento promedio de variabilidad entre cualquiera de las Edades de los Estudiantes Universitarios objeto de estudio y la Mx es de 3,63 años. Existen algunas distancias mucho mayores, así como distancias mucho menores, con respecto de la Media Aritmética, sin embargo en promedio es aproximadamente 3,63 años. Dado que resulta difícil evaluar la magnitud de este valor, es necesario determinar la Variabilidad Relativa a través de un Coeficiente Estadístico que se estudiará la próxima clase.

DESVIACIÓN TÍPICA O ESTÁNDAR MUESTRAL (Sx)*:* es la raíz cuadrada de las desviaciones cuadráticas de los valores de la variable, con respecto a la media aritmética, en otras palabras es la raíz cuadrada de la varianza (Armas, 1988). Esta medida de variabilidad determina el grado de dispersión de los datos en investigación.

$$Sx = \sqrt{S^2x}$$

¿Cómo comparar dos o más distribuciones de datos? Cuando se desean comparar la variabilidad de dos o más conjuntos de datos, **NO** se puede efectuar con la **Desviación Típica** porque los datos para cada distribución van arrojar resultados que **NO** se pueden comparar por las unidades de medida de la (s) magnitud (es) de las dos series de datos objeto de estudio, para dar solución a esta problemática se estudiarán los *Coeficientes de Variación*, por ser el más ampliamente utilizado.

Las Medidas de Tendencia Central y de Variabilidad dan información sobre una muestra, resulta oportuno preguntarse **¿tiene sentido usar estas magnitudes para comparar dos series de datos?**

Ejemplo si solicitan comparar la **variabilidad** de las **Estaturas (metros)** de una muestra de Estudiantes de Quinto Año pertenecientes a dos Instituciones Educativas de países diferentes **¿Dará información útil?**

¿Qué ocurre si se compara la Estatura (metros) de una muestra de Estudiantes de primer año de la Escuela de Medicina con respecto a su Edad (años cumplidos)? Tanto la Media Aritmética como la Desviación Típica, **NO** pueden comparar las variables antes mencionadas porque poseen unidades de medida diferente, por lo tanto, **NO** tiene ningún sentido.

El **problema no deriva** sólo de que una de las magnitudes sea *longitud* y la otra sea *tiempo*. El mismo **problema** se plantea si medimos **dos escalas de medida diferentes**, en otras palabras, múltiplos y submúltiplos de una misma unidad de medida. **Por ejemplo:** la **Estatura** de una muestra de la Institución "X" medida en **Kilómetros**, con una muestra de la Institución "Y" medida en **centímetros.**

El **problema no se resuelve** tomando las mismas escalas para ambas muestras. **Por ejemplo**: Medir las **Masas** (Toneladas) de unas hormigas con unos elefantes, lo lógico es que la dispersión de la variable Masa de las hormigas sea prácticamente nula.

En los dos primeros casos mencionados anteriormente, el problema viene de las unidades de medida de la (s) magnitud (es) de las dos series de datos objeto de estudio, y en el tercero de la diferencia enorme entre las medias de ambas muestras. El **Coeficiente de Variación (CV)** permite evitar estos problemas, pues elimina las unidades de medida de la (s) magnitud (es) de las dos series de datos objeto de estudio de las variables y no considera las **Mx** de las muestras, por lo tanto, se convierte en el Coeficiente de Variación Relativa más confiable.

El **Coeficiente de Variación** (**CV**) propuesto por **Karl Pearson en 1895** con el propósito de contrastar la variabilidad entre varias distribuciones de frecuencias (Martínez, 2008; p.199). Este coeficiente es de gran difusión en los libros de texto, páginas web y de mayor utilización en la actualidad.

¿Cómo se define el CV? Martínez (2008, p.199), lo define como el cociente entre la Desviación Típica (Sx) y la Media Aritmética (Mx), el mismo autor efectúa una distinción del Coeficiente de Variación:

Es **Relativo** cuando se divide la Desviación Típica (Sx) entre la Media Aritmética (Mx), es decir:

$$CV = \frac{Sx}{Mx}, Mx > 0 \text{ (Martínez, 2008; p.199)}$$

Es **Porcentual** cuando se divide la Desviación Típica (Sx) entre la Media Aritmética (Mx) y posteriormente se multiplica por 100, es decir:

$$CV = \frac{Sx}{Mx} \times 100, Mx > 0 \text{ (Martínez, 2008; p.199)}$$

Para interpretarlo se usará la siguiente tabla:

Considerando la siguiente Tabla 6:

Tabla 6. Interpretación cualitativa del CV.

Es importante recordar que el **corchete** es un símbolo que incluye el extremo inferior y/o superior del intervalo, mientras que el **paréntesis** es un símbolo que excluye el extremo inferior y/o superior del intervalo **por ejemplo**: en el intervalo (0,00 a 0,25) no se incluyen los valores 0,00 y 0,25, en cambio en el intervalo [0,75 a 1,00) se incluye el valor 0,75 y no se incluye el valor 1,00.

Valor del CV	Interpretación cualitativa
0,00	Nula
(0, 00 a 0,25)	Muy Baja
[0,25 a 0,50)	Moderadamente Baja
[0,50 a 0,75)	Moderadamente Alta
[0,75 a 1,00)	Muy Alta
1,00	Máxima

Nota: una distribución es considerada **Relativamente Normal** cuando el CV oscila entre 0,45 (moderadamente baja) y 0,55 (moderadamente alta).

Se deben comparar el resultado del **Coeficiente de Variación** de las variables en estudio y **el resultado menor determinará la distribución de menor variabilidad.**

Por otro lado, cabe señalar que se considera que la Media Aritmética es un **valor "esperado"** o el valor de la Media esperada en un número infinito de veces, debido a que si se toman varias muestras al azar de la misma población (conjunto de datos), bajo condiciones idénticas el mismo valor tiende a ser obtenido, *con cierto margen de error por exceso o por defecto*, dependiendo de la magnitud de la variabilidad observada en las observaciones de cada muestra, la cual depende de la variabilidad de la *población total*. El valor específico del error proveniente del valor "esperado" **se estima** mediante el **Error Típico de la Media (ETM)**, el cual se calcula dividiendo la Desviación Típica entre la Raíz Cuadrada del tamaño de la muestra:

$$ETM = \frac{Sx}{\sqrt{n}} \text{ (Hays, 1973; p. 285)}$$

La **Desviación Típica (Sx)**, el **Error Típico de la Media (ETM)** y el **Coeficiente de Variación (CV)** nos indican el grado de estabilidad de las medidas de tendencia central, por lo cual resulta fundamental que las tres medidas antes mencionadas sean presentadas en forma conjunta.

¿Regla empírica? lleva este nombre porque está basada en la experiencia y está sujeta a comprobación y comparación de los datos en estudio con los valores de probabilidad teóricos. Los valores obtenidos en la media, la mediana y la moda deben coincidir, es decir debe ser simétrica, además de unimodal (poseer una sola moda).

Está caracterizada por dos parámetros: la media aritmética y la desviación típica, donde *Geométricamente,* se interpreta la media aritmética como un factor de traslación y la desviación típica como un factor de escala o grado de dispersión y *probabilísticamente, s*i se toma intervalos centrados en la media aritmética, y cuyos extremos están una distancia de la desviación típica, entonces tenemos una probabilidad de tomar el 68% de los datos en estudio; si los extremos están a una distancia de dos desviaciones típicas, entonces hay una probabilidad de tomar el 95% de los elementos de la distribución y si los extremos están a distancia de tres desviaciones típicas, entonces existe una probabilidad de tomar el 99% de los datos que se están investigando.

Referencias

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

Chipia, J. (2013). *Efectividad de un Programa de Enseñanza/Aprendizaje sobre Estadística Descriptiva utilizado Calc de Open.* Memoria en extenso (N°45) del VII CIBEM, 16-20 de septiembre de 2013.

Glass, G. y Stanley, J. (1986). *Métodos estadísticos aplicados a las ciencias sociales*. México: Prentice Hall.

Hays, W. L. (1973). *Statistics for the social sciences* (2da. Ed.). New York: Holt, Rinehart and Wilson.

Martínez, C. (2008). *Estadística y muestreo* (12da. Ed.). Bogotá: ECOE Ediciones.

CAPÍTULO 6. MEDIDAS DE TENDENCIA NO CENTRAL

Estas medidas descriptivas permiten ubicar la posición que ocupa un valor dentro de un conjunto de datos, se calcula para variables de tipo cualitativo ordinal y de tipo cuantitativo (discreta y continua), cabe agregar que los resultados se expresan en las mismas unidades de los datos en estudio.

PERCENTILES: son aquellos valores que dividen los datos ordenados en cien partes iguales. Existen noventa y nueve percentiles, dicha medida hace referencia a un porcentaje de casos por debajo del percentil y otro porcentaje por encima. Entre dos percentiles consecutivos cualesquiera se encuentra un 1% o 1/100 partes de los datos.

Ejemplo. *Calcule e interprete ¿Cuál es la edad que deja el 70% de las edades por debajo?*
Datos de la edad (meses cumplidos) 205, 196, 195, 198, 180, 206.

Los percentiles para datos no agrupados se calculan así:

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

Siendo:

n : Total de la muestra.

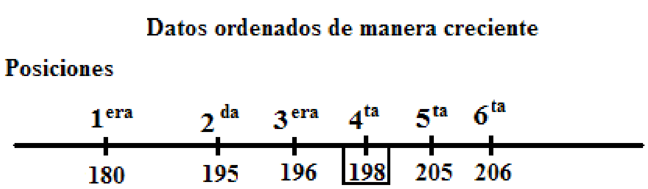
k : Percentil a calcular.

Para el ejemplo se debe calcular el percentil 70 (P_{70}).

Primero se ordenan los datos de manera creciente. Datos ordenados: 180, 195, 196, 198, 205, 206.
Luego multiplicamos la cantidad de datos por el percentil a determinar y el resultado se divide entre cien, esto se realiza con el fin de obtener la posición:

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

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



Entonces 4 es la posición del P_{70} **=198** meses. **Interpretación:** 198 meses es la edad que deja el 70% de las edades por debajo y el 30% por encima.

Ejemplo. *Calcule e interprete ¿Cuál es la edad que deja el 80% de las edades por encima?*
Datos de la edad (meses cumplidos) 205, 196, 195, 198, 180, 206.

Para el ejemplo se debe calcular el percentil 20 (P_{20}).
Primero se ordenan los datos de manera creciente. Datos ordenados: 180, 195, 196, 198, 205, 206. Luego multiplicamos la cantidad de datos por el percentil a determinar y el resultado se divide entre cien, esto se realiza con el fin de obtener la posición:

$$P_{20} = \frac{6x20}{100} = 1{,}2 \approx 1$$

Entonces 1 es la posición del **P_{70} =180** meses

Interpretación: 180 meses es la edad que deja el 20% de las edades por debajo y el 80% por encima.

DECILES: son valores que dividen los datos ordenados en diez partes iguales. Existen nueve deciles, dicha medida deja un porcentaje de datos por debajo del decil y otro porcentaje por encima. Entre dos deciles consecutivos cualesquiera se encuentra un 10% o 1/10 partes de los elementos.

Ejemplo: Calcule e interprete ¿Cuál es la edad que deja el 20% de las edades por encima?
Datos de edad (meses): 205, 196, 195, 198, 180, 206.

Los deciles para datos no agrupados, se calculan así:

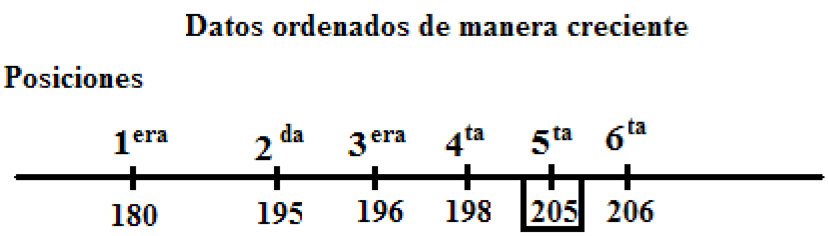
$$D_k = \frac{n \times k}{10}.$$

Siendo:
n : Total de la muestra.
k : Decil a calcular.

Para el ejemplo se debe calcular el decil 8 (D_8)
Primero se ordenan los datos de manera creciente. Datos ordenados: 180, 195, 196, 198, 205, 206. Luego se multiplica la cantidad de datos por el decil a determinar y el resultado se divide entre diez, esto se realiza con el fin de obtener la posición:

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

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



Entonces 5 es la posición del D_8 **=205** meses

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

CUARTILES: son valores que dividen los datos ordenados en cuatro partes iguales. Existen tres cuartiles, por lo tanto dicha medida hace referencia a un porcentaje de casos por debajo del cuartil y otro porcentaje por encima Entre dos percentiles consecutivos cualesquiera se encuentra un 25% o 1/4 partes de los datos.

Ejemplo: *Calcule e interprete ¿Cuál es el valor que toma como máximo el 25% de las edades de los estudiantes?* Datos de edad (meses): 205, 196, 195, 198, 180, 206.

Los cuartiles para datos no agrupados se calculan así:

$$Q_k = \frac{n \times k}{4}.$$

Siendo:

n : Total de la muestra.

k : Cuartil a calcular.

Para el ejemplo se debe calcular el cuartil 1 (Q_1)

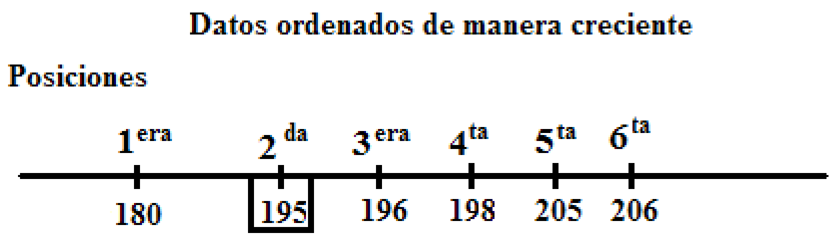
Primero se ordenan los datos de manera creciente:

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

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

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

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



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

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

Referencias

Armas, J. (1988). *Estadística sencilla descriptiva.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Chipia, J. y Lara, C. (2008). *Módulo para la enseñanza–aprendizaje de la estadística en el bachillerato mediante la solución de situaciones problemas.* Mérida: Tesis de Grado mención Publicación en la Universidad de Los Andes.

CAPÍTULO 7. PROBABILIDAD ELEMENTAL

En el lenguaje cotidiano se escuchan expresiones que hacen referencia a la probabilidad, tales como:

- Probablemente visite a María el fin de semana.
- Es muy probable que los Leones del Caracas gane el próximo domingo.
- Al lanzar un dado, es más probable que salga 1,2, 3, o 4 a que salga 4 o 6.
- Es casi seguro que el empleo de médico lo tome una mujer en lugar de un hombre.

Las expresiones anteriores muestran una idea intuitiva del concepto de **Probabilidad,** dichas ideas reflejan la posibilidad de ocurrencia de hechos o sucesos, en los cuales está presente la incertidumbre en cuanto a lo que puede acontecer.

Se define la **PROBABILIDAD**, como la teoría que cuantifica los posibles resultados de un experimento en el cual está presente la incertidumbre o aleatoriedad. En otras palabras, se habla de **Probabilidad** cuando en un evento intervienen procesos físicos, biológicos o sociales que generan observaciones, y cuyo resultado no es posible predecir con exactitud.

La **Probabilidad** intuitivamente, se traduce como un número que va a reflejar la posibilidad de que algo ocurra bajo ciertas condiciones. Cuando nos referimos a la Probabilidad, estamos hablando de la *ocurrencia de cierto evento, simple o compuesto, en un* **experimento aleatorio***.* Si denotamos por E a un evento, entonces vamos a representar por P (E) a la probabilidad de que el **evento** E ocurra.

Cuando se señalan los **EXPERIMENTOS** se hace de la manera más amplia posible, es decir, no sólo incluyen situaciones asociados a escenarios experimentales en un laboratorio, sino también se contemplan cualesquiera otras situaciones que den origen a sucesos de interés. Los experimentos en Probabilidad pueden ser determinísticos o aleatorio. En la cuantificación de la **Probabilidad** de un evento, son importantes el **tipo de experimento** (*simple:* por un solo punto muestral; *compuesto:* si contiene más de un punto muestral).

EXPERIMENTOS DETERMINÍSTICOS: También llamados exactos, los cuales se caracterizan porque cada vez que se realizan bajo condiciones similares, producen el mismo resultado. Estos fenómenos no son de interés para la Estadística.

En general a la *ciencia Estadística*, y en particular a la *teoría de la probabilidad*, les interesa y fundamentan su desarrollo y aplicación en los denominados **experimentos aleatorios**.

EXPERIMENTO ALEATORIO: Es cualquier acción o proceso que no se tiene certeza de su resultado final, hasta tanto no se ejecute. Este tipo de experimento debe satisfacer con los siguientes requerimientos:

- Puede repetirse un número ilimitado de veces bajo las mismas condiciones.
- Es posible conocer por adelantado todos los posibles resultados a que puede dar origen.
- No puede predecirse con exactitud un resultado en una realización particular del experimento.

Ejemplo: Si se desea formar un equipo de voleibol con 5 jugadores, el nombre de los seleccionados no se sabrá con certeza hasta que no se realicen las pruebas correspondientes y se elijan a los 5 deportistas. Se puede conocer la lista de todos los deportistas inscritos, pero no la lista de los seleccionados.

ESPACIO MUESTRAL (S): De un experimento aleatorio, es el conjunto de todos los posibles resultados al realizar el experimento. **Ejemplo:** el espacio muestral de lanzar un dado una vez es
$S = \{1, 2, 3, 4, 5, 6\}$.

EVENTOS (E): Es un subconjunto del espacio muestral. **Ejemplo:** al lanzar un dado una vez, determine el evento que en la cara superior aparece un número par, el evento es $E = \{2, 4, 6\}$.

Resulta oportuno señalar:

- Si dos o más eventos **son mutuamente excluyentes**, entonces la intersección de los eventos es vacío.
- Si dos o más eventos **son exhaustivos**, entonces la unión de los eventos es el espacio muestral.
- Si dos o más eventos son exhaustivos y a su vez son mutuamente excluyentes, entonces son **colectivamente exhaustivos.**

Ejercicio. Hallar el espacio muestral del siguiente experimento: El papá de un bebé próximo a nacer quiere que su hijo se llame Juan, Camilo o Felipe. La mamá por su parte, pretende que se llame Andrés o Pablo. Para que ambos queden felices deciden combinar los nombres propuestos, considerando que primero irá el del papá y, luego, el de la mamá ¿De cuántas formas diferentes se pueden proponer un nombre para el bebé?

Solución:

El espacio muestral serán todas las combinaciones que se puedan armar con los 3 nombres que propone el papá y los 2 que propone la mamá; se debe tener en cuenta que primero irá el del papá y luego el de la madre. Por lo tanto, tenemos:

S = {Juan Andrés, Juan Pablo, Camilo Andrés, Camilo Pablo, Felipe Andrés, Felipe Pablo}

Ejercicio. Hallar el espacio muestral del siguiente experimento: Los candidatos para formar la nueva junta del consejo comunal son Carlos, Josefa, Elías y Marina. Se requiere que la junta esté compuesta por un presidente y un secretario ¿De cuántas formas se puede formar esta junta?

Solución:

Sean: C=Carlos, J=Josefa, E=Elías, M=Marina

En el espacio muestral se debe considerar el orden en que se seleccione la junta.

S = {(C,J), (J,C), (C,E), (E,C), (C,M), (M,C), (J,E), (E,J), (J,M), (M,J), (E.M), (M,E)}

PROBABILIDAD DE OCURRENCIA DE UN EVENTO

- Criterio de frecuencia relativa
- Criterio de probabilidad subjetiva
- Criterio de equiprobabilidad

CRITERIO DE FRECUENCIA RELATIVA: Es la proporción de veces que ocurre A en “n” repeticiones del experimento.

$$P(A) = \frac{n_A}{n}$$

Este concepto de probabilidad es de carácter empírico. Cabe agregar que el cociente tiende a estabilizarse, a medida que se incrementa indefinidamente el número de repeticiones. A lo anterior se le denomina propiedad de estabilización de la frecuencia relativa de un evento, o ley de regularidad estadística de un experimento aleatorio.

Críticas del criterio de frecuencia relativa:

- Sólo es aplicable a experimentos que pueden repetirse un número ilimitado de veces.
- ¿Qué tan grande tiene que ser n?
- No es posible garantizar que para todos los posibles eventos asociados al experimento, el valor tienda a estabilizarse.
- En la práctica es difícil asegurar que el experimento se va a repetir, bajo las mismas condiciones.

CRITERIO DE PROBABILIDAD SUBJETIVA: El cual establece que la probabilidad es un hecho subjetivo y expresa el grado de creencia o convicción personal (también llamada "corazonadas") que se tiene la ocurrencia de un evento o suceso, fundamentado en las evidencias o en el sentir que se tenga en particular cuando se realiza el experimento.

CRITERIO DE EQUIPROBABILIDAD: Sea un experimento aleatorio finito, con k posibilidades, donde cada una tiene la misma posibilidad de ocurrir. Entonces la probabilidad de ocurrencia de cada punto muestral es 1/k y si A es cualquier evento con r puntos muestrales, se tiene:

$$P(A) = \frac{r}{k} = \frac{\#A}{\#S}$$

Caracterización de la equiprobabilidad

- Un espacio muestral bajo las características anteriores se dice que es equiprobable.
- Se aplica cuando se asume simetría o razones especiales inherentes al experimento.
- En estos casos la probabilidad se determina a priori y no es necesario repetir el experimento.
- Es aplicable a espacios muestrales finitos e infinitos.

DEFINICIÓN AXIOMÁTICA DE PROBABILIDAD

a) $P(A) \geq 0$

b) $P(S) = 1$

c) $Si\ A_1, A_2, \ldots es\ una\ colección\ infinita\ \ de\ eventos\ mutuamente\ excluyentes, se\ cumple$:

$$P\left(\bigcup_{i=1}^{\infty} Ai\right) = \sum_{i=1}^{\infty} P(Ai)$$

TEOREMAS BÁSICOS DE PROBABILIDAD

Teorema 1. $P(\emptyset) = 0$

Teorema 2. Si A y B son eventos mutuamente excluyentes, se cumple $P(A \cup B) = P(A) + P(B)$

Teorema 3. Si A y B son eventos cualesquiera, se cumple $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Teorema 4. $P(A^c) = 1 - P(A)$

Teorema 5. Para cualquier evento A, se cumple $0 \leq P(A) \leq 1$

Teorema 6. Si A y B son eventos cualesquiera, se cumple $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Teorema 7. Si A y B son eventos tales que A está contenido en B, entonces $P(A) \leq P(B)$

Ejemplo. Se usan 7 fichas numeradas del 1 al 7 en una caja y se seleccionan dos de ellas de manera simultánea. Calcular la probabilidad de los siguientes eventos:

El espacio muestral de este experimento aleatorio es:

$$S = \begin{Bmatrix} (1,2),(1,3,),(1,4),(1,5),(1,6),(1,7), \\ (2,3),(2,4),(2,5),(2,6),(2,7), \\ (3,4),(3,5),(3,6),(3,7),(4,5), \\ (4,6),(4,7),(5,6),(5,7),(6,7) \end{Bmatrix}$$

a) La suma de las dos fichas es 7.

A=La suma de las dos fichas es 7

A = {(1,6), (2,5), (3,4)}

#S=21

#A=3

$$P(A) = \frac{\#A}{\#S} = \frac{3}{21} = 0{,}1429$$

Interpretación: 0,1429 es la probabilidad de que la suma de las dos fichas sea 7.

b) La suma de las dos fichas es menor que 14.

Sea **B** el evento que consiste en que la suma de los dos números sea menor a 14. Se tiene que **S=B**, por tanto, es un evento *seguro,* o su *probabilidad es 1.*

c) El número mayor de las 2 fichas seleccionadas es 2.

C = El número mayor de las 2 fichas seleccionadas es 2.

C = {(1,2)}

#C=1

$$P(C) = \frac{\#C}{\#S} = \frac{1}{21} = 0{,}0476$$

d) Las dos fichas seleccionadas tengan el mismo número.

Este resultado no es posible o vacío y por teorema 1, $P(\emptyset) = 0$ se concluye que la Probabilidad es cero.

PROBABILIDAD Y TABLAS DE CONTINGENCIA: A partir de la información suministrada en una tabla de doble entrada, de contingencia o bivariable, es posible calcular las probabilidades de algunos eventos asociados con las variables presentadas.

Ejercicio. A partir de los datos de la Tabla 7, calcule:

Tabla 7. Tipo de café y Procedencia de los asistentes a un Congreso de Salud.
Mérida, Venezuela. Marzo, 2015

		TIPO DE CAFÉ			
		Negro	**Marrón**	**Con leche**	**Total**
PROCEDENCIA	**Europa**	13	11	8	32
	América	12	2	4	18
	Total	25	13	12	50

Fuente: Datos supuestos, Marzo de 2015.

1. ¿Cuál es la probabilidad de que una persona proceda de Europa o consuma café negro?

Para dar solución se debe hallar la unión de los eventos, utilizando el teorema 3:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Se van a considerar los eventos:

A: la persona procede de Europa.

B: la persona consume café negro.

Realizando el procedimiento por partes se tiene:

$$P(A) = \frac{\#A}{\#S} = \frac{32}{50} = 0{,}64 \qquad P(B) = \frac{\#A}{\#S} = \frac{25}{50} = 0{,}5 \qquad P(A \cap B) = \frac{\#A \cap B}{\#S} = \frac{13}{50} = 0{,}26$$

Sustituyendo los resultados:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0{,}64 + 0{,}5 - 0{,}26 = 0{,}88$$

En conclusión 0,88 es la Probabilidad de de que una persona proceda de Europa o consuma café negro.

Tabla 8. Tipo de incapacidad de los usuarios en cada Centro de Salud.

Mérida, Venezuela. Marzo, 2015

		TIPO DE INCAPACIDAD			
		Brazos	**Piernas**	**General**	**Total**
CENTRO DE SALUD	**IVSS**	50	82	15	147
	"Sor Juana Inés de la Cruz"	35	45	25	105
	IAHULA	75	90	38	203
	Total	160	217	78	455

Fuente: Datos supuestos, Marzo de 2015.

Si se selecciona un paciente al azar de los que se incluyeron en el reporte

a) ¿Cuál es la probabilidad de que su incapacidad no sea en las piernas?

Sean **A:** usuarios con incapacidad en las piernas y **A^c:** usuarios sin incapacidad en las piernas.

Si $P(A) = \frac{217}{455} = 0{,}4769$ Para dar respuesta a la pregunta se debe calcular:

$P(A^c) = 1 - P(A) = 1 - 0{,}4769 = 0{,}5231$

b) ¿Cuál es la probabilidad de que provenga del IVSS o del IAHULA?

Sean **B:** el usuario proviene del IVSS y **C:** el usuario proviene del IAHULA.

Dado que $B \cap C = \emptyset$ Se debe calcular $P(B \cup C) = P(B) + P(C) = \frac{147}{455} + \frac{203}{455} = \frac{350}{455} = 0{,}7692$

c) ¿Cuál es la probabilidad de que provenga del IAHULA y tenga incapacidad en sus brazos?

Sean **C:** el usuario proviene del IAHULA y **D:** el usuario tiene incapacidad en los brazos.

Se debe calcular: $P(C \cap D) = \frac{75}{455} = 0{,}1648$

Tabla 9. Sexo y Facultad donde presentaron los aspirantes a ingresar a la ULA.

Mérida, Venezuela. Marzo, 2015.

		FACULTADES				
		Med.	**Der.**	**Hum.**	**Cien.**	**Total**
SEXO	**Masculino**	0,055	0,120	0,085	0,212	0,472
	Femenino	0,163	0,201	0,116	0,049	0,528
	Total	0,218	0,321	0,201	0,26	1

Fuente: Datos supuestos, Marzo de 2015.

Si se selecciona a un aspirante al azar:

a) ¿Cuál es la probabilidad de que no se haya presentado a la facultad de medicina o sea de sexo femenino?

Sean E: el aspirante presentó en la facultad de medicina y F: el aspirante es de sexo femenino.

Se halla: $P(E^c \cup F) = P(E^c) + P(F) - P(E^c \cap F)$

$P(E^c \cup F) = (1 - 0{,}218) + (0{,}528) - (0{,}201 + 0{,}116 + 0{,}049) = 0{,}944$

b) ¿Cuál es la probabilidad de que no se haya presentado a la facultad de ciencias o no sea de sexo femenino?

Sean G: el aspirante presentó en la facultad de ciencias y H: el aspirante es de sexo femenino.

Se calcula: $P(G^c \cup H^c) = P(G^c) + P(H^c) - P(G^c \cap H^c)$

$P(G^c \cup H^c) = (1 - 0{,}26) + (1 - 0{,}528) - (0{,}055 + 0{,}120 + 0{,}085) = 0{,}952$

PROBABILIDAD CONDICIONAL: En ocasiones es necesario calcular la probabilidad de evento sujeto a alguna condición o algún otro evento que ya ocurrió, **por ejemplo**, si se considera el ejercicio en el cual hay una vía principal en la que existen dos semáforos. Al pasar un vehículo por la vía se pueden definir dos eventos:

A: el auto se detenga en el primer semáforo.
B: el auto se detenga en el segundo semáforo.

Es posible considerar la probabilidad de que el vehículo se deba detener en el segundo semáforo, sabiendo que se detuvo en el primero. En este caso, se establece una condición que está representada por el evento A; es decir, se quiere reconocer la probabilidad de que ocurra el evento B, sabiendo que ya ocurrió el evento A.

A partir de la idea intuitiva anteriormente señalada, se tiene que la **PROBABILIDAD CONDICIONAL**, para dos eventos A y B, en el cual uno de ellos hace el papel de condición sobre el otro, es decir, se debe considerar que uno de los dos eventos ya ocurrió y se quiere saber qué sucede con el otro. Por lo tanto, se define que dados los eventos A y B, se llama probabilidad condicional de A dado B o P(A/B) así:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

También se puede ver la probabilidad condicional de B dado A o P (B/A) así:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}, P(A) \neq 0$$

Ejemplo: Para conformar el comité de emergencias de los estudiantes de medicina, se presentaron 4 candidatos. Luisa (L), Marina (M), Carlos (C) y Eleonora (E). Se decidió que se haría la primera votación entre los miembros de la comisión y quien ganara sería el presidente. Luego, se haría una nueva votación entre los 3 candidatos que quedaran y el de mayor votación sería el coordinador. Si se sabe que el presidente fue Carlos **¿cuál es la probabilidad de que Marina sea la coordinadora?**

Sean A: Carlos es el presidente del comité y B: Marina es la coordinadora del comité.

A es la condición sobre B, por lo tanto, se debe calcular la probabilidad de B dado A **P(B/A)**

$$A \cap B = \{(C, M)\} \quad P(A \cap B) = \frac{1}{12}$$

$$A = \{(C, L), (C, M), (C, E)\} \quad P(A) = \frac{3}{12}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{12}}{\frac{3}{12}} = \frac{1}{3} = 0{,}3333$$

Ejemplo. En un concesionario funciona una oficina de ventas de seguros para automóviles. Se ha determinado que la probabilidad de que un cliente compre su automóvil en el concesionario es de 0,58. La probabilidad de que un cliente compre una póliza de seguro contra todo riesgo es de 0,42 y la probabilidad de que el cliente compre el auto y el seguro en el concesionario es de 0,35. Si llega un cliente al concesionario y compra el auto, **¿cuál es la probabilidad de que compre el seguro allí?**

Sean E: el cliente compra el automóvil en el concesionario y F: el cliente compra el seguro en el concesionario.

$$P(E \cap F) = 0{,}35$$

$$P(E) = 0{,}58$$

$$P(F/E) = \frac{P(E \cap F)}{P(E)} = \frac{0{,}35}{0{,}58} = 0{,}6034$$

EVENTOS INDEPENDIENTES: Dos sucesos aleatorios son independientes entre sí, cuando la probabilidad de cada uno de ellos, no está influida porque el otro suceso ocurra o no, es decir, cuando ambos sucesos no están relacionados.

Si dos eventos son independientes, se cumple cualquiera de las siguientes sentencias:

$$i)\ P(A/B) = P(A), \qquad P(B) \neq 0$$

$$ii)\ P(B/A) = P(B), \qquad P(A) \neq 0$$

$$iii)\ P(A \cap B) = P(A).P(B);\ P(A) \neq 0, P(B) \neq 0$$

Dos eventos no son independientes si al menos una de las sentencias no se cumple. Es importante aclarar que los términos independiente y mutuamente excluyente no significan la misma cosa.

Ejemplo. En un grupo de pacientes, que consta de 60 mujeres y 40 varones, se observa que 24 mujeres y 16 varones usan lentes. Si un paciente es elegido aleatoriamente, la probabilidad de que el paciente utilice lentes es 0,4.

a) ¿Cuál es la probabilidad de que un paciente elegido aleatoriamente use lentes dado que es varón?

Sea **A:** el paciente es varón y **B:** el paciente usa lentes.

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{16/_{100}}{40/_{100}} = \frac{16}{40} = 0{,}4$$

Podemos observar que $P(B/A) = P(B)$

Se puede concluir que los eventos ser varón y usar lentes, en el grupo en estudio son independientes.

b) Determine si usar lentes y no ser varón son eventos independientes.

$\boldsymbol{A^C}$**:** el paciente no es varón.
B: el paciente usa lentes.

$$P(B/A^C) = \frac{P(A^C \cap B)}{P(A^C)} = \frac{24/_{100}}{60/_{100}} = \frac{24}{60} = 0{,}4$$

Podemos observar que $P(B/A^C) = P(B)$

TEOREMA DE BAYES: Es una aplicación de las probabilidades condicionales, que son llamadas probabilidades a posteriori, en el campo de ciencias de la salud, se utiliza ampliamente la aplicación de leyes de probabilidad, y conceptos relacionados en la evaluación de pruebas de detección y criterios de diagnóstico. A los médicos les interesa tener mayor capacidad para predecir correctamente la presencia o ausencia de una enfermedad, a partir del conocimiento de los resultados (positivos o negativos) de pruebas de detección y el estado de los síntomas (presentes o ausentes) que se manifiestan.

El teorema de Bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A. Es decir que sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Este sencillo ejemplo la alta relevancia del teorema en cuestión para la ciencia en todas sus ramas, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados.

En pruebas de detección, se debe considerar con cuidado que no siempre son pruebas infalibles, en otras palabras, el procedimiento puede arrojar un *falso positivo* o un *falso negativo*. Un **falso positivo** resulta cuando la prueba indica que el estado es positivo, cuando en realidad es negativo. Un **falso negativo** resulta cuando una prueba indica que un estado es negativo, cuando en realidad es positivo.

El Teorema sirve para evaluar la utilidad de los resultados de la prueba y el estado de los síntomas, por lo tanto determina si el individuo tiene o no alguna enfermedad, para lo cual se debe responder a las siguientes interrogantes:

1. Dado que un individuo tiene la enfermedad, ¿qué probabilidad existe de que la prueba resulte positiva (o la presencia de un síntoma)?
2. Dado que un individuo no tiene la enfermedad, ¿cuál es la probabilidad de que la prueba resulte negativa (o la ausencia de un síntoma)?
3. Dada una prueba positiva de detección (o la presencia de un síntoma), ¿qué probabilidad existe de que el individuo tenga la enfermedad?
4. Dado el resultado negativo de una prueba de detección (o la ausencia de un síntoma), ¿cuál es la probabilidad de que el individuo no tenga la enfermedad?

Para dar respuesta a las interrogantes planteadas se muestra la Tabla 10, la cual considera una muestra de *n* individuos (con *n* muy grande) clasificados en referencia cruzada según el estado de enfermedad y el resultado del estado de detección, como se tiene que "a" es el número de individuos que tienen la enfermedad con un resultado positivo en la prueba de detección.

Tabla 10. Presencia de enfermedad y resultado de la prueba.

		Enfermedad		
		Presente (D)	**Ausente** (D^C)	**Total**
Resultado de prueba	**Positivo** (T)	a	b	a+b
	Negativo (T^C)	c	d	c+d
	Total	a+c	b+d	n

SENSIBILIDAD de una prueba o síntoma: es la probabilidad de un resultado positivo de la prueba dada la presencia de la enfermedad, se calcula: $P(T/D) = \frac{a}{a+c}$

ESPECIFICIDAD de una prueba o síntoma: es la probabilidad de un resultado negativo de la prueba dada la ausencia de la enfermedad, se calcula: $P(T^c/D^c) = \frac{d}{b+d}$

POSITIVIDAD de una prueba o síntoma: es la probabilidad de que un individuo tenga la enfermedad, dado que el resultado positivo en la prueba de detección, se calcula $P(D/T)$ de la siguiente manera:

$$P(D/T) = \frac{P(T/D)P(D)}{P(T/D)P(D) + P(T/D^c)P(D^c)}$$

El **numerador** de la ecuación, es igual a la sensibilidad por la tasa (de prevalencia) de la enfermedad. El **denominador** de la ecuación, es igual a la sensibilidad por la tasa de prevalencia más el término 1 menos la sensibilidad por el termino 1 menos la tasa de la enfermedad.

Se puede simplificar utilizando la siguiente fórmula:

$$P(D/T) = \frac{a}{a+b}$$

NEGATIVIDAD de una prueba o síntoma: es la probabilidad de que un individuo tenga la enfermedad, dado que el resultado positivo en la prueba de detección, se calcula $P(D^C/T^C)$ de la siguiente manera:

$$P(D^c/T^c) = \frac{P(T^c/D^c)P(D^c)}{P(T^c/D^c)P(D^c) + P(T^c/D)P(D)}$$

Se puede simplificar de la siguiente manera:

$$P(D^c/T^c) = \frac{d}{c+d}$$

Ejemplo: Un equipo de investigación pretende evaluar una prueba de detección propuesta para la enfermedad de Alzheimer. La prueba se basa en una muestra aleatoria de 950 pacientes. Los datos se obtuvieron de una población de individuos con edades de 65 años o más y se muestran en la Tabla 11.

Tabla 11. Presencia de Alzheimer y resultado de la prueba.

		Presencia de Alzheimer		
		Si (D)	**No** (D^C)	**Total**
Resultado de prueba	**Positivo** (T)	416	25	441
	Negativo (T^C)	34	475	509
	Total	450	500	950

Cálculo e interpretación de:

SENSIBILIDAD

$$P(T/D) = \frac{a}{a+c} = \frac{416}{450} = 0{,}9244$$

Interpretación: 0,9244 es la probabilidad de que la prueba sea positivo, dada la presencia de la enfermedad.

ESPECIFICIDAD

$$P(T^c/D^C) = \frac{d}{b+d} = \frac{475}{500} = 0{,}95$$

Interpretación: 0,95 es la probabilidad de que la prueba sea negativa, dada la ausencia de la enfermedad.

POSITIVIDAD

$$P(D/T) = \frac{a}{a+b} = \frac{416}{441} = 0{,}9433$$

Interpretación: 0,9433 es la probabilidad de que un individuo tenga la enfermedad, dado que presenta un resultado positivo en la prueba.

NEGATIVIDAD

$$P(D^c/T^c) = \frac{d}{c+d} = \frac{475}{509} = 0{,}9332$$

Interpretación: 0,9332 es la probabilidad de que un sujeto no tenga la enfermedad, dado que presenta un resultado negativo en la prueba.

Referencias

Armas, J. (1992). *Estadística sencilla: probabilidades.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Daniel, W. (2010). *Bioestadística: Base para el análisis de las ciencias de la salud (*4a. Ed.). México: Limusa Wiley.

CAPÍTULO 8. DISTRIBUCIONES DE PROBABILIDAD

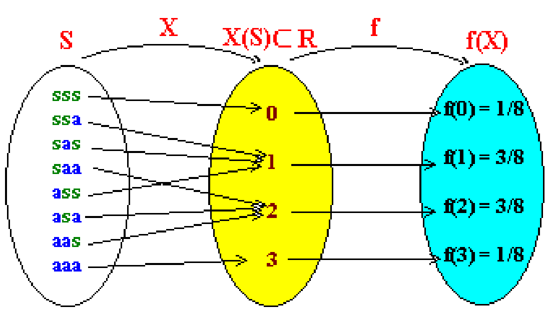
En muchos estudios no se desea saber cuál evento ocurrió, sino el número de veces que ha ocurrido un evento. Por ejemplo, al lanzar dos monedas, podríamos estar interesados en el número de caras que ocurrieron. Al nacer 5 niños, quisiéramos saber cuántos son varones. Los ejemplos anteriores tienen la característica de que a cada uno de los elementos del espacio muestral se le asigna un número real, que indica el número de veces que está presente el evento de interés. Dicha asignación se realiza a través de una función la cual se denomina **variable aleatoria**.

VARIABLE ALEATORIA: Es una función que asigna un número real, a cada resultado del espacio muestral, de un experimento aleatorio. En otras palabras, es una función X definida:

$$\boldsymbol{X}: \boldsymbol{S} \rightarrow \boldsymbol{R}$$

Por tanto, es una **función** cuyo *dominio es el espacio muestral* y el *rango es el conjunto de los número reales*.

El **espacio muestral** en muchas ocasiones, no está constituido por números, pero a través de la **variable aleatoria**, se puede expresar en forma numérica todo tipo de espacio muestral, lo cual facilita el **análisis** de sus aspectos más relevantes. Gráficamente se puede observar por medio del siguiente diagrama sagital:



La **distribución de probabilidad** de una variable aleatoria refleja su comportamiento probabilístico.

NOTA: En muchos casos ocurre que los elementos del espacio muestral también son números, entonces, X queda definida por X (w) = w (función identidad), es una variable aleatoria. En dicha situación, el mismo experimento aleatorio define una variable aleatoria, con dominio y rango iguales.

Ejemplo. Se lanza una moneda.

S = {C, S}.

Sea X = {Número de caras}.

Esta función asigna los siguientes valores a los elementos del espacio muestral:

Si es cara, w = C, entonces, X (w) = 1.

Si es sello, w = S, entonces, X (w) = 0.

Por lo tanto, la **variable aleatoria** X toma los valores {0, 1}

Ejemplo. Se lanzan dos monedas.

S= {CC, CS, SC, SS}.

X = {Número de caras}

Esta función asigna los siguientes valores a los elementos del espacio muestral:

Si w = CC, entonces, X (w) = 2.

Si w = CS, entonces, X (w) = 1.

Si w = SC, entonces, X (w) = 1.

Si w = SS, entonces, X (w) = 0.

Por lo tanto, la **variable aleatoria** X toma los valores: {0, 1, 2}

Ejemplo. Se lanza un dado

S= {1, 2, 3, 4, 5, 6}.

X = {Número de la cara superior del dado}

Es decir, X (w)=w, entonces X es una variable aleatoria.

El dominio y el rango de X es el conjunto {1, 2, 3, 4, 5, 6}.

Ejemplo. Secuencia del sexo de los dos primeros recién nacidos en un Hospital.

Si utilizamos M para masculino y F para femenino:

S= {MM, MF, FM, FF}.

X = {Número de femenino en los dos recién nacidos}

Si w = MM, entonces, X (w) = 2.

Si w = MF = FM, entonces, X (w) = 1.

Si w = FF, entonces, X (w) = 0.

El **dominio** de X es el conjunto {MM, MF, FM, FF} y el **rango** es el conjunto {0,1, 2}.

¿CÓMO DETERMINAR LAS PROBABILIDADES DE "S" A TRAVÉS DE "X"?

La probabilidad de ocurrencia de este ejemplo de cada elemento de S, es igual a 1/4

X = {Número de féminas en los dos nacimientos}

X puede tomar: 2 para MM, 1 para MF y FM y 0 para FF.

Se tiene: P(X=0) = P (FF) = 1/4.

P(X=1) = P (MF, FM) = P (MF) + P (FM) = 1/4 + 1/4 = 1/2

P(X=2) = P (MM) = 1/4

LAS VARIABLES ALEATORIAS SE DIVIDEN EN DISCRETAS Y CONTINUAS:

Son **discretas** cuando puede tomar un número finito o infinito numerable de valores.

Ejemplos.

- El número de accidentes de tránsito que ocurren en una autopista en un lapso de tiempo determinado.
- El número de artículos defectuosos que se encuentran en una muestra aleatoria de 20 artículos producidos por una máquina.
- El número de veces que se lanza una moneda hasta que salga la primera cara.
- El número de hermanos de una persona seleccionada al azar.

Son **continuas** cuando puede tomar un número infinito no numerable de valores. Alternativamente, se puede definir como aquella variable que puede tomar cualquier valor dentro de un intervalo de números reales.

Ejemplos.

- El tiempo de espera de un paciente antes de ser atendido.
- La edad, la estatura, el peso, la presión arterial, la temperatura.
- Ingresos y egresos de una familia.

DISTRIBUCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA DISCRETA "X": La Variable Aleatoria discreta X es el conjunto formado por los valores x que puede tomar esa variable y las correspondientes probabilidades P (X=x). En toda distribución de probabilidad de una V. A. discreta, debe cumplirse que todas las probabilidades tienen que estar comprendidas entre 0 y 1 y la suma de ellas, es igual a la unidad (1).

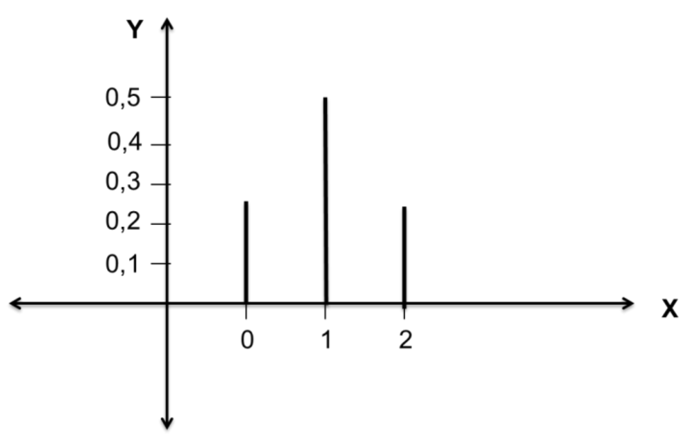
La distribución de probabilidad del ejemplo en el cual se lanzan dos monedas quedaría de la siguiente manera:

x	**0**	**1**	**2**
P (X=x)	¼	1/2	¼

Es importante notar que, la suma de las probabilidades es 1. Aquellos valores que no toma la variable, tiene una probabilidad igual a cero (0).

DIAGRAMA DE LÍNEAS: Este diagrama sirve para representar una distribución de probabilidad discreta. Se construye colocando sobre el eje de las abscisas (**eje x**) los **valores de la variable** y sobre el eje de las ordenadas (**eje y**) las **probabilidades**. Luego se traza para cada valor de la variable una línea paralela al eje de las ordenadas, cuya altura es igual a la correspondiente probabilidad asociada a ese valor.

Gráfico 11. Distribución de probabilidad del experimento se lanzan dos monedas.



FUNCIÓN DE DISTRIBUCIÓN DE UNA VARIABLE ALEATORIA DISCRETA: Es la función que la probabilidad de X sea menor o igual que un valor determinado x, siendo x cualquier número real, es decir:

$$F_x(x) = P(X \leq x) = \sum_{t \leq x} P(X = t), \forall x \in R$$

A esta función también se le conoce con el nombre de distribución de probabilidad acumulativa.

Ejemplo. Consideremos una variable aleatoria con distribución de probabilidad, dada por:

x	**1**	**2**	**3**	**4**
P (X=x)	0,2	0,3	0,3	0,2

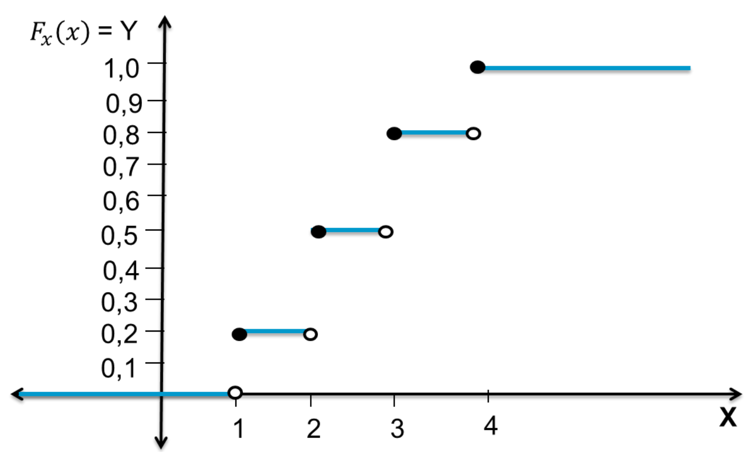
$$F_x(x) \begin{cases} 0 & x < 1 \\ 0{,}2 & 1 \leq x < 2 \\ 0{,}5 & 2 \leq x < 3 \\ 0{,}8 & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

A partir de $F_x(x)$ se pueden calcular probabilidades tales como

$P(X = 3) = F_x(3) - F_x(2) = 0{,}8 - 0{,}5 = 0{,}3$

REPRESENTACIÓN GRAFICA DE $F_X(x)$: Es un gráfico en forma escalonada, se construye colocando en el eje y a $F_X(x)$ y en el eje x, los valores de x, quedando de la siguiente manera:

Gráfico 12. Función De Distribución De Una Variable Aleatoria Discreta.



ESPERANZA MATEMÁTICA E(x): También llamada esperanza, valor esperado, media poblacional (µ) o media de una variable aleatoria, es el número que formaliza la idea de valor medio de un fenómeno aleatorio. Cuando la variable aleatoria es discreta, la esperanza es igual a la suma de la probabilidad de cada posible suceso aleatorio, multiplicado por el valor de dicho suceso. La esperanza matemática representa la cantidad media que se "espera“, como resultado de un experimento aleatorio, cuando la probabilidad de cada suceso se mantiene constante y el experimento se repite un elevado número de veces. Se calcula así:

$$E(x) = \sum x_i . P(X = x)$$

Ejemplo. Halle la esperanza matemática de variable aleatoria con distribución de probabilidad, dada por:

x	**1**	**2**	**3**	**4**
P (X=x)	0,2	0,3	0,3	0,2

$E(x) = \sum x_i \times P(X = x) = (1 \times 0{,}2) + (2 \times 0{,}3) + (3 \times 0{,}3) + (4 \times 0{,}2)$

$E(x) = 0{,}2 + 0{,}6 + 0{,}9 + 0{,}8 = 2{,}5$

Ejemplo. Si una persona compra un ticket en una rifa, en la que puede ganar en el primer premio Bs. 5000, segundo premio Bs. 2000 con probabilidades de: 0.001 y 0.003 respectivamente ¿Cuál sería el precio justo a pagar por el ticket?

$E(x) = (5000 \times 0{,}001) + (2000 \times 0{,}003) = 11\ Bs.$

Ejemplo. Determinar la **esperanza matemática** del juego y si es favorable, En un juego de 2 monedas, **Gana** Bs. 1 si aparece una cara, Gana Bs. 2 si aparecen dos caras y **Pierde** Bs. 5 si no aparece cara.

$$S= \{CC, CS, SC, SS\}. \quad \begin{cases} p(+1) = 2/4 \\ p(+2) = 1/4 \\ p(-5) = 1/4 \end{cases}$$

$E(x) = \left(1 \times \frac{2}{4}\right) + \left(2 \times \frac{1}{4}\right) - \left(5 \times \frac{1}{4}\right) = -\frac{1}{4}$

El juego es desfavorable y el resultado indica que por cada Bs. 4 que juegue pierde Bs. 1.

VARIANZA MATEMÁTICA V(x) o σ^2: De una variable aleatoria es una medida de dispersión, que explica la desviación cuadrada de los valores de dicha variable, con respecto a su media, multiplicada por la probabilidad de ocurrencia del evento. Esta medida se expresa en unidades cuadráticas o elevadas al cuadrado. La varianza tiene como valor mínimo cero (0). La varianza puede verse muy influida por los valores atípicos. En tales casos se recomienda el uso de otras medidas de dispersión más robustas. Se calcula con la siguiente fórmula:

$$V(x) = \sum (x_i - E(x)^2)P(X = x)$$

DESVIACIÓN ESTÁNDAR DE(x) o σ: Es la raíz cuadrada de la varianza, es una medida que sirve para determinar la dispersión de la variable objeto de análisis, expresada en unidades de medida lineales. Se calcula con la siguiente fórmula:

$$DE(x) = \sqrt{V(x)}$$

Ejemplo. En una fábrica de insumos médicos se ha determinado que el número de insumos defectuosos producidos en cada turno de trabajo, es una variable aleatoria con la siguiente distribución de probabilidad:

x	0	1	2	3	4
P (X=x)	0,90	0,06	0,02	0,01	0,01

Hallar la esperanza, varianza y desviación estándar de esa distribución de probabilidad.

Esperanza Matemática

$$E(x) = [(0x0{,}9) + (1x0{,}06) + (2x0{,}02) + (3x0{,}01) + (4x0{,}01)] = 0{,}17\ insumos$$

Varianza Matemática

$$V(x) = [(0 - 0{,}17)^2 x0{,}90] + [(1 - 0{,}17)^2 x0{,}06] + [(2 - 0{,}17)^2 x0{,}02] + [(3 - 0{,}17)^2 x0{,}01] + [(4 - 0{,}17)^2 x0{,}01] = 0{,}3611\ insumos^2$$

$$V(x) = 0{,}3611\ insumos^2$$

Desviación estándar

$$DE(x) = \sqrt{0{,}3611\ insumos^2} = 0{,}6009\ insumos$$

DISTRIBUCIÓN DE PROBABILIDAD: Es un modelo de probabilidad de una variable aleatoria que permite la representación teórica simplificada de un fenómeno real y la elaboración de afirmaciones probabilísticas sobre ese fenómeno. Lo anterior facilita la toma de decisiones y la previsión de hechos que pueden ocurrir.

El estudio de las probabilidades se fundamentan en el estudio de experimentos aleatorios, es decir, en situaciones reales en las cuales existe incertidumbre y variabilidad sobre lo que pueda ocurrir, aún si ésta situación se repite muchas veces en condiciones similares. Un modelo para representar teóricamente la realidad en términos probabilísticos, lo constituye una variable aleatoria (V. A.) y su correspondiente distribución de probabilidad, porque la primera refleja la situación particular y la segunda expresa el comportamiento de la variable.

Resulta oportuno señalar que cualquier modelo de probabilidad que se intente aplicar construirá una aproximación a la realidad, tendrá sus limitantes e idealizaciones y se basará en supuestos que requieren de cumplimiento. A través de inferencia estadística se decide en base a informaciones muestrales, si es razonable asumir que un determinado modelo de probabilidad, es el que mejor describe el comportamiento de una población.

Cuando el problema que se analiza puede representarse mediante una variable aleatoria discreta, el correspondiente modelo de probabilidad se categoriza como **DISCRETO**, el cual se define, *como el conjunto de valores que puede tomar una variable discreta acompañado de sus respectivas probabilidades.*

EXPERIMENTO DE BERNOULLI: Es un experimento en el cual sólo hay dos posibles resultados que llamaremos Éxito (E) y Fracaso (F). La denominación anterior se usará de manera convencional, depende del interés y naturaleza del fenómeno estudiado. Por lo general, cualquier experimento aleatorio puede considerarse como experimento de Bernoulli redefiniéndolo. No necesariamente se debe asociar éxito a cosas agradables y fracaso a aspectos desagradables, *ejemplo:* al realizar una biopsia el paciente considera el resultado negativo como un éxito y el positivo como un fracaso.

Ejemplos.

- Registrar el resultado de un examen de un estudiante, en términos de aprobar o reprobar.
- Aprobación o negación de una solicitud de crédito ante un banco.
- Determinar la calidad de un producto en las opciones de defectuoso o no defectuoso.
- Determinar si un caso sospechoso, es positivo o negativo.
- Medir la estatura de una persona y registrar si mide igual y meno de 1,60 metros, o no.
- Observar el color de ojos de una persona seleccionada al azar fijándose, si son negros o no.
- Seleccionar al azar un automóvil en una ciudad y registrar, si es de fabricación italiana o no.
- Lanzar un dado y registrar si sale par o no.

Resulta oportuno indicar que el experimento de Bernoulli de la variable aleatoria X, se define como el número de éxitos, dicha variable sólo puede tomar los valores 0 y 1, denotándose por p la probabilidad de éxito y por 1-p la probabilidad de fracaso.

DISTRIBUCIÓN DE PROBABILIDAD DE BERNOULLI: Esta distribución depende solamente del valor de probabilidad de éxito **p** y en consecuencia, para diferentes valores de **p** se obtienen diferentes distribuciones de Bernoulli, a este valor **p** se le denomina parámetro de la distribución.

Viene dada por:

$$P(X=x)\begin{cases} 1-p & \text{si } x=0 \\ p & \text{si } x=1 \\ 0 & \text{en otro caso} \end{cases}$$

Alternativamente se puede expresar con la siguiente fórmula:

$P(X=x) = p^x(1-p)^{1-x}, x=0,1$

EXPERIMENTO BINOMIAL: Es aquel que cumple con las siguientes propiedades

- Consiste en **n ensayos de Bernoulli**, es decir, la repetición n veces de un experimento que consta de dos posibles resultados, que llamaremos éxito o fracaso.

- La probabilidad **p de éxito se mantiene constante** en cada uno de los n ensayos de Bernoulli y en consecuencia la probabilidad de **fracaso 1-p, también se mantiene constante** en cada una de las pruebas o ensayos.
- Los n ensayos de Bernoulli **son independientes entre sí**, o sea que el resultado de un ensayo no afecta el resultado de los demás.

Ejemplos.

- Lanzar 10 veces una moneda.
- Responder al azar (sin pensar) una prueba de verdadero o falso, constituida por 20 preguntas.
- Inspeccionar 8 artículos seleccionados al azar, producidos en una fábrica, para determinar los defectuosos y los no defectuosos.
- Examinar 7 casos sospechosos, para determinar los positivos y los negativos.
- Otros experimentos con n-repeticiones en los cuales se seleccionen aleatoriamente una muestra de una población dicotómica.

Ejemplo. Selección de una muestra con reposición.

Una población está conformada por 16 personas de las cuales 4 son hombres y 12 son mujeres. Se desea seleccionar una muestra aleatoria con reposición de 3 personas y registrar el sexo de cada una. Si denominamos éxito, al hecho de que una persona sea del sexo femenino, entonces la situación anterior constituye un experimento binomial, el cual arroja como probabilidad:

$$p = P(éxito) = \frac{12}{16} = 0{,}75$$

Ejemplo. Selección de una muestra sin reposición.

Consideremos la situación del ejemplo anterior, pero en éste ejemplo se tomará la muestra sin reposición. No es un experimento binomial, por cuanto los ensayos no son independientes entre sí.

Ejemplo. Selección de una muestra sin reposición cuando la población es grande.

Consideremos el caso de seleccionar una muestra aleatoria de 3 personas, de una población constituida por 200 hombres y 1000 mujeres. Estrictamente no es un experimento binomial, porque los ensayos no son independientes. Pero desde el punto de vista práctico, se puede aceptar debido a que, cuando la muestra es grande los valores en la muestra con reposición y sin reposición son muy aproximados.

DISTRIBUCIÓN BINOMIAL. Dado un experimento binomial cualquiera, vamos a definir la variable aleatoria X, como el número de éxitos que se obtienen en los n ensayos de Bernoulli; a esta variable se le denomina variable binomial y su correspondiente distribución de probabilidad, se conoce como distribución binomial.

La variable aleatoria X que representa el número de éxitos en n ensayos de un experimento binomial, sigue una distribución dada por:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0,1,2, \dots, n$$

La cual se denomina **Distribución Binomial**, con n y p como parámetros y para diferentes valores se obtienen distintas distribuciones binomiales. En algunos textos se define la distribución binomial, con la probabilidad de fracaso por q=1-p, quedando expresada de la siguiente manera:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0,1,2, \dots, n \;\; Siendo\, p + q = 1$$

Nota: el nombre de la distribución binomial proviene del hecho de que las **PROBABILIDADES BINOMIALES** para $x = 0, 1, \dots, n$ son precisamente los términos o sumandos del desarrollo del binomio $(q + p)^n$, también llamado **Binomio de Newton**.

Ejemplo. Se tiene una tasa de prevalencia del 30% de asma en los habitantes de una ciudad. Determinar en una muestra aleatoria de 4 personas la probabilidad de que: a) ninguna padezca de asma; b) más de dos sufran de asma.

Solución. Sea X el número de personas que sufren de asma en una muestra de 4 personas, es decir: X (n = 4; p = 0,3)

a) $P(X = 0) = \binom{4}{0} (0{,}3)^0 (0{,}7)^4 = 0{,}2401$

b) $P(X \geq 3) = \binom{4}{3} (0{,}3)^3 (0{,}7)^1 + \binom{4}{4} (0{,}3)^4 (0{,}7)^0 = 0{,}0756 + 0{,}0081 = 0{,}0837$

Ejemplo. Si 9 de cada 10 recién nacidos de una región son niños, determinar la probabilidad de que en una muestra aleatoria de 12 recién nacidos, exista un máximo de 10 niños. Si **X** representa el número de **niños** y **Y** el número de **niñas** en el grupo de 12 recién nacidos, tenemos que:

$$X\ (n = 12, p = 0{,}9)$$
$$Y\ (n = 12, p = 0{,}1)$$

Para dar respuesta se puede usar X o Y, veamos los dos casos:

Para X:

$$P(X \leq 10) = 1 - P(X \geq 11) = 1 - \binom{12}{11}(0{,}9)^{11}(0{,}1)^{1} - \binom{12}{12}(0{,}9)^{12}(0{,}1)^{0}$$
$$P(X \leq 10) = 1 - 0{,}37657 - 0{,}28242 = 0{,}34101$$

Para Y:

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - \binom{12}{1}(0{,}1)^{1}(0{,}9)^{11} - \binom{12}{0}(0{,}1)^{1}(0{,}9)^{12}$$
$$P(Y \geq 2) = 1 - 0{,}37657 - 0{,}28242 = 0{,}34101$$

Ejemplo. Uso de la Tabla de Distribución Binomial.

Si se quiere hallar el valor para n=4, x=0, p=0,45, se pude hacer así:

$$P(X = 0) = \binom{4}{0}(0{,}45)^{0}(0{,}55)^{4} = 0{,}0915$$

También se puede utilizar la Tabla que sigue, tal como se muestra:

PROBABILIDADES BINOMIALES

n	x	p									
		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	0	0,9500	0,9000	0,8500	0,8000	0,7500	0,7000	0,6500	0,6000	0,5500	0,5000
1	1	0,0500	0,1000	0,1500	0,2000	0,2500	0,3000	0,3500	0,4000	0,4500	0,5000
2	0	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500
2	1	0,0950	0,1800	0,2550	0,3200	0,3750	0,4200	0,4550	0,4800	0,4950	0,5000
2	2	0,0025	0,0100	0,0225	0,0400	0,0625	0,0900	0,1225	0,1600	0,2025	0,2500
3	0	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250
3	1	0,1354	0,2430	0,3251	0,3840	0,4219	0,4410	0,4436	0,4320	0,4084	0,3750
3	2	0,0071	0,0270	0,0574	0,0960	0,1406	0,1890	0,2389	0,2880	0,3341	0,3750
3	3	0,0001	0,0010	0,0034	0,0080	0,0156	0,0270	0,0429	0,0640	0,0911	0,1250
4	0	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625
4	1	0,1715	0,2916	0,3685	0,4096	0,4219	0,4116	0,3845	0,3456	0,2995	0,2500
4	2	0,0135	0,0486	0,0975	0,1536	0,2109	0,2646	0,3105	0,3456	0,3675	0,3750
4	3	0,0005	0,0036	0,0115	0,0256	0,0469	0,0756	0,1115	0,1536	0,2005	0,2500
4	4	0,0000	0,0001	0,0005	0,0016	0,0039	0,0081	0,0150	0,0256	0,0410	0,0625

DISTRIBUCIÓN NORMAL: También llamada distribución gaussiana o de Gauss, es un modelo continuo de probabilidad. Es la distribución más conocida, más aplicada y de mayor desarrollo teórico entre las diferentes distribuciones de probabilidad, incluyendo discretas y continuas, por tanto, mantiene un lugar de primacía entre los modelos de probabilidad.

Una variable aleatoria $X(\mu, \sigma)$ sigue una **DISTRIBUCIÓN NORMAL**, si su función de densidad viene dada por:

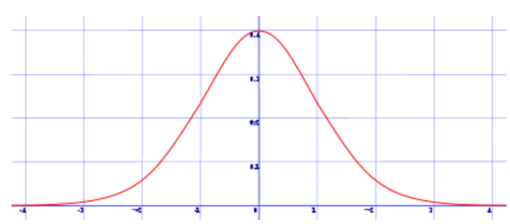
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

$$Siendo\ \mu\ y\ \sigma los\ parámetros,\ \mu\ y\ \sigma\ \epsilon\ R,\ \sigma > 0$$

$$\pi = 3{,}14159\ldots$$

$$e = 2{,}71828\ldots$$

GRÁFICO DE LA DISTRIBUCIÓN NORMAL: Se conoce como curva normal y presenta las siguientes características



- Tiene forma acampanada.
- Es simétrica con respecto a la µ y unimodal, es decir, µ=Md=Mo.
- Los extremos de la curva o colas son asintóticas con respecto al eje X.
- El área bajo la curva es igual a uno (1).
- Curtosis y Asimetría igual a cero (0).

INTERPRETACIÓN GOMÉTRICA:

- El parámetro μ es un factor de *traslación*, por lo tanto, si aumenta μ se traslada a la derecha, en el caso contrario se traslada a la izquierda.
- El parámetro σ es un factor de *dispersión*, por lo tanto, si aumenta σ se acumulan los datos en el centro de la distribución, en el caso opuesto se extienden a lo largo del eje X.

INTERPRETACIÓN PROBABILÍSTICA: Cumple las siguientes igualdades

$P(\mu - \sigma < x < \mu + \sigma) = 0{,}6826$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0{,}9545$

$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0{,}9978$

DISTRIBUCIÓN NORMAL ESTÁNDARIZADA: Es toda distribución normal con μ = 0 y $\sigma^2 = 1$, se hace un proceso de transformación, que se denomina estandarización o tipificación. Se acostumbra a designar Z (0; 1), a toda variable que siga una distribución normal estandarizada, siendo

$$Z = \frac{X - \mu}{\sigma}$$

En la Distribución Normal Estandarizada se cumple:

$P(-1 < Z < 1) = 0{,}6826$

$P(-2 < Z < 2) = 0{,}9545$

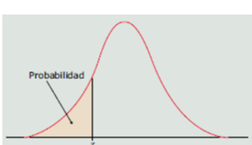
$P(-3 < Z < 3) = 0{,}9978$

Para calcular probabilidades asociadas a una variable normal estandarizada con la tabla correspondiente, se procede de la siguiente manera:

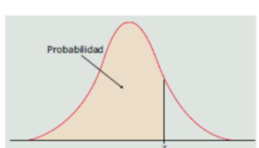
i) Las probabilidades del tipo $P(Z \leq z)$ se buscan directamente en la tabla.

ii) Las probabilidades del tipo $P(Z \geq z) = 1 - P(Z \leq z)$.

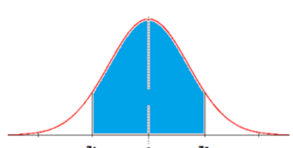
iii) Las probabilidades de la forma: $P(z_1 \leq Z \leq z_2) = P(Z \leq z_2) - P(Z \leq z_1)$



Caso i)



Caso ii)



Caso iii)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Ejemplo. Determinar los valores de probabilidad a partir de la tabla de distribución normal estandarizada

$i)\ P(Z < -1{,}21)$

$ii)\ P(Z > 2{,}43)$

$iii)\ P(-1{,}33 \leq Z < 2{,}2)$

Para i) se cruza 1,2 con 0,01, dando como resultado el valor de probabilidad de 0,1131, por lo tanto:

$$P(Z < -1{,}2) = 0{,}1131$$

Para ii) se cruza 2,4 con 0,03 dando 0,9925, posteriormente se realiza lo siguiente:

$$P(Z > 2{,}43) = 1 - P(Z < 2{,}43) = 1 - 0{,}9925 = 0{,}0075$$

Para iii) se hallan los valores por separado en la tabla como sigue

$$P(-1{,}33 < Z < 2{,}2)$$

$$P(Z < 2{,}2) = 0{,}9861$$

$$P(Z < -1{,}33) = 0{,}0916$$

Luego se realiza el siguiente procedimiento:

$$P(-1{,}33 < Z < 2{,}2) = P(Z < 2{,}2) - P(Z < -1{,}33) = 0{,}9861 - 0{,}0916 = 0{,}8943$$

Problema. El ingreso diario de un consultorio médico es de μ=30000 Bs y σ=5000 Bs. Determine: a) la probabilidad de que en un día cualquiera los ingresos sean superiores a 45000 Bs b) Si el consultorio registra pérdidas cuando el ingreso diario es inferior a 20000 Bs, calcule el porcentaje de días durante el año que se espera que el consultorio presente pérdidas.

Solución a):

$$Z = \{ingreso\ diario\ del\ consultorio\ médico\}$$
$$Z(\mu = 30000; \sigma = 5000)$$
$$a)\ P(Z > 45000) = 1 - P\left(Z < \frac{45000 - 30000}{5000}\right)$$
$$P(Z > 45000) = 1 - P(Z < 3) = 1 - 0{,}9987 = 0{,}0013$$

Lo que indica un evento poco probable.

Solución b):

$$P(Z < 20000) = P\left(Z < \frac{20000 - 30000}{5000}\right)$$
$$P(Z < 20000) = P(Z < -2) = 0{,}0228$$
$$Luego,\ \%_{Z<-2} = 0{,}0228\ x\ 100 = 2{,}28\%$$

En consecuencia 2,28% de los días del año, el consultorio médico registrará pérdidas.

DISTRIBUCIÓN CHI CUADRADO: surge como una distribución de la suma de los cuadrados de las variables aleatorias independientes normalmente distribuidas. Esta distribución dentro de sus aplicaciones, sirve para hacer inferencia acerca de una varianza de una población; de prueba de bondad de ajuste para distribuciones de probabilidad; prueba de independencia, para determinar si dos atributos son independientes entre sí (Ovalles y Moret, 2001).

Si $X_1, X_2, \dots, X_n$ son n variables aleatorias normales independientes, de μ=0 y σ^2=1, la variable definida como:

$$Y_n = X_1^2 + \cdots + X_n^2 = \sum_{i=1}^{n} X_i^2$$

Una variable aleatoria continúa X, tiene una distribución Chi-cuadrado (X^2) con n grados de libertad, si su función de densidad está dada por:

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)\sqrt{2^n}} x^{(n-2)/2} e^{-x/2} \qquad x > 0$$

Características:

- El parámetro de la distribución Chi-cuadrado, es el número de grados de libertad.
- Es una variable aleatoria continua.
- Es una distribución unimodal y asimétrica positiva cuando los grados de libertad son pequeños.
- Su campo de variación se encuentra en el intervalo (0, +∞), debido a que es una suma de cuadrados
- Presenta una familia de distribuciones Chi-cuadrado, es decir, para cada grado de libertad, existe una distribución Chi-cuadrado.
- A medida que aumentan los grados de libertad, la distribución Chi-cuadrado, se aproxima a una distribución normal.

Ejemplo. Uso de la tabla de Distribución Chi-cuadrado

$i)P(X^2 \leq 0{,}30)\ con\ 8\ grados\ de\ libertad = \mathbf{9{,}52}$

$ii)P(X^2 \geq 0{,}05)\ con\ 6\ grados\ de\ libertad = \mathbf{1{,}63}$

$iii)P(0{,}1 \leq X^2 \leq 0{,}9) con\ 5\ grados\ de\ libertad = \mathbf{9{,}24} \leq X^2 \leq \mathbf{1{,}61}$

Grados de libertad	Probabilidad										
	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,01	0,001
1	0,004	0,02	0,06	0,15	0,46	1,07	1,64	2,71	3,84	6,64	10,83
2	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99	9,21	13,82
3	0,35	0,58	1,01	1,42	2,37	3,66	4,64	6,25	7,82	11,34	16,27
4	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	13,28	18,47
5	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,07	15,09	20,52
6	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59	16,81	22,46
7	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	14,07	18,48	24,32
8	2,73	3,49	4,59	5.53	7,34	9,52	11,03	13,36	15,51	20,09	26,12
9	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92	21,67	27,88
10	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31	23,21	29,59

Interpretación de los resultados

$i)P(X^2 \leq 0{,}30)\ con\ 8\ grados\ de\ libertad = 9{,}52.$

9,52 es el valor de X^2 que deja a su izquierda, un área de 0,30 o 30% en una distribución con 8 grados de libertad.

$ii)\ P(X^2 \geq 0{,}05) con\ 6\ grados\ de\ libertad = 1{,}63 = P(X^2 \leq 0{,}95) = 1{,}63$

1,63 es el valor de X^2 que deja a su derecha, un área de 0,05 o 5% en una distribución con 6 grados de libertad.

$iii) P(0{,}1 \leq X^2 \leq 0{,}9) con\ 5\ grados\ de\ libertad = 9{,}24 \leq X^2 \leq 1{,}61$

1,61 es el valor de X^2 que deja a su izquierda un área de 0,1 o 10%, mientras que el 9,24 deja a la derecha un área de 0,1 o 10%, en una distribución con 5 grados de libertad, en otras palabras el 80% de la variable X^2se encuentra entre 1,61 y 9,24.

DISTRIBUCIÓN T DE STUDENT: Surge como el cociente de dos variables aleatorias independientes: una normal estandarizada (Z) en el numerador y la raíz cuadrada de la Chi-cuadrada (X^2) entre sus grados de libertad (n) en el denominador, es decir:

$$t = \frac{Z}{\sqrt{\frac{X^2}{n}}}$$

La función de densidad con n grados de libertad es:

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)^2} \quad -\infty \quad < x < \quad +\infty$$

CARACTERÍSTICAS:

- El parámetro de la distribución t, es el número de grados de libertad.
- Es una variable aleatoria continua.
- Es una distribución unimodal y simétrica con respecto a su media.
- Su campo de variación está en el intervalo $(-\infty, +\infty)$
- Existe una distribución t diferente para cada grado de libertad, con μ=0 y σ diferente dependiendo de los grados de libertad.
- A medida que aumentan los grados de libertad, la distribución t-student se aproxima a una distribución normal.

Ejemplo. Hallar el valor de probabilidad asociado en la tabla de la Distribución t de Student

i) Hallar el área derecha de t, con n=12 y α=0,05.

Grados de libertad	Probabilidades							
	0,40	0,25	0,15	0,10	0,05	0,025	0,01	0,005
1	0,3249	1,0000	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,2887	0,8165	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,2767	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,2707	0,7407	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,2672	0,7267	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,2648	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,2632	0,7111	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,2619	0,7064	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,2610	0,7027	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,2602	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,2596	0,6974	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,2590	0,6955	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,2586	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,2582	0,6924	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,2579	0,6912	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467
16	0,2576	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,2573	0,6892	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,2571	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,2569	0,6876	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,2567	0,6870	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453
21	0,2566	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314
22	0,2564	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188

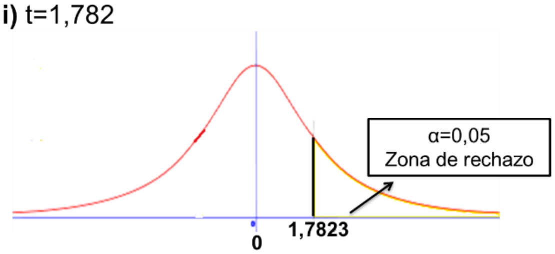
Se determina en la tabla cruzando los datos suministrados, por lo tanto el valor es t=1,7823

ii) Hallar el área izquierda de t, con n=12 y α=0,05.

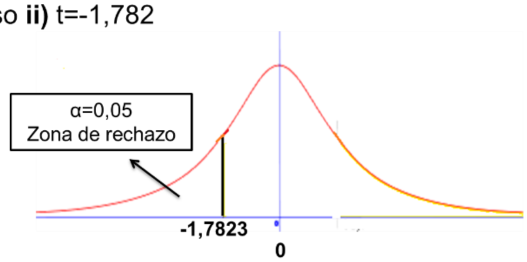
La distribución t de Student es simétrica, por ello el valor t es el mismo de i) pero con signo contrario, es decir, t=-1,7823.

Interpretación geométrica del resultado:

Caso **i)** t=1,782



Caso **ii)** t=-1,782

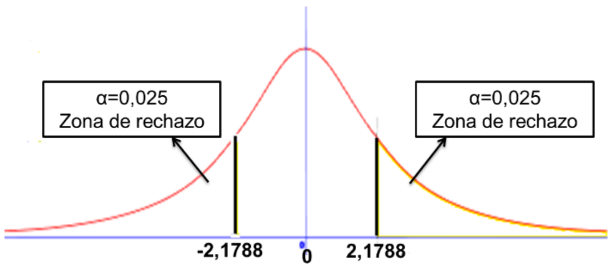


Ejemplo. Dada una variable t de Student con 12 grados de libertad. Hallar el valor t que deja un área total de 5% en ambos extremos.

Lo primero que se debe calcular es $\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$

Usando la tabla α=0,025 se tiene que **t=2,1788** para la *cola derecha* y debido a la simetría de la distribución de t, el valor de la *cola izquierda* es el mismo pero con signo cambiado, es decir: **t=-21788**

Interpretación geométrica:



Referencias

Armas, J. (1992). *Estadística sencilla: probabilidades.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Ovalles, A. y Moret, C. (2001). *Manual de estadística ii*. Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

CAPÍTULO 9. PRUEBAS DE HIPÓTESIS ESTADÍSTICAS

Los procedimientos de inferencia estadística permiten sacar conclusiones de una población, usando la información aportada por una muestra aleatoria representativa, presentando los resultados un pequeño margen de error. Es necesario tener presente que en la inferencia estadística se pueden hacer estimaciones y contrastes de hipótesis.

ESTIMACIÓN: Procedimiento mediante el cual se estima el valor de un parámetro poblacional (los parámetros poblacionales son estimados a partir de estadísticos), un estadístico es denominado **estimador** cuando se usa para estimar un parámetro. **Ejemplo:** estimar la proporción de estudiantes universitarios que fuman o el número de horas diarias que dedican al estudio semanalmente.

Asociado a cada parámetro poblacional se pueden encontrar uno o varios estimadores. No todo estimador es un buen estimador. Por ello, de entre todos los estimadores asociados a un parámetro poblacional, se escoge al mejor estimador del parámetro poblacional.

¿Qué condiciones debe cumplir un estimador para ser considerado el mejor estimador? Son cuatro las condiciones que se exigen: Ausencia de sesgo (imparcialidad), consistencia, eficacia y suficiencia.

Ausencia de sesgo o imparcialidad de un estimador se presenta cuando los valores obtenidos para el estimador se centran alrededor del parámetro poblacional. Es decir, la media de la distribución del estimador es igual al parámetro poblacional.

Eficiencia de un estimador imparcial A se dice eficiente en comparación con otro B, si la varianza de A es menor que la varianza de B

Consistente: es aquel estimador que al aumentar el tamaño de la muestra, converge en probabilidad al parámetro que estima.

Suficiente: cuando incluye toda la información que la muestra puede proporcionar acerca del parámetro.

Existen dos tipos de estimaciones:

Estimación Puntual: Procedimiento mediante el cual se estima el valor puntual de un parámetro poblacional. Se le dice puntual ya que se obtiene como resultado un valor numérico para el parámetro poblacional. **Ejemplo:** se estima que un 35% de los estudiantes de cierta universidad fuman.

Estimación por Intervalos: Procedimiento mediante el cual se estima el valor de un parámetro poblacional usando un intervalo numérico; acá el resultado obtenido es un intervalo dentro del cual se espera, con cierto grado de confianza, se encuentre el verdadero valor del parámetro poblacional.

CONTRASTE DE HIPÓTESIS: Procedimiento empleado para decidir si una hipótesis hecha sobre una población debe ser rechazada o mantenida. Los contrastes surgen al someter a contraste hipótesis de investigación. **Ejemplo:** determinar si los estudiantes universitarios dedican en promedio 8 horas semanales al estudio.

PASOS PARA EL CONTRASTE DE HIPÓTESIS

PASO 1. Hipótesis Estadísticas

La hipótesis son los supuestos sujetos a contrastación y en una investigación cuantitativa que plantee hipótesis, se pueden probar mediante un contraste de hipótesis estadísticas.

Hipótesis nula (H_0): Es la hipótesis que se formula con la esperanza de rechazarla, por lo general se redacta en negativo. Puede especificar: (1) que un parámetro es igual a un valor, (2) que dos o más parámetros poblacionales son iguales o (3) que la población se distribuye según cierta forma.

Hipótesis alternativa (H_1): Es la hipótesis que contradice lo especificado en la hipótesis nula. Generalmente, la hipótesis alternativa coincide con alguna parte o la totalidad de la hipótesis de investigación propuesta.

Planteamientos de hipótesis estadísticas:

Un planteamiento de hipótesis se denomina de **una cola**, cuando la hipótesis alternativa es unilateral. **Ejemplo:**

$$H_0: \mu = a$$
$$H_1: \mu > a \text{ ó } H_1: \mu < a$$

Un planteamiento de hipótesis se denomina de **dos colas**, cuando la hipótesis alternativa es bilateral. **Ejemplo:**

$$H_0: \mu = a$$
$$H_1: \mu \neq a$$

PASO 2. Nivel de Significancia

Es la Probabilidad que se asume como error aleatorio, porque al rechazar o aceptar la hipótesis nula propuesta existe la posibilidad de cometer un error.

Existen dos tipos de error posibles: Error tipo I y Error tipo II.

Error tipo I: ocurre cuando se rechaza la hipótesis nula siendo verdadera. A la probabilidad de ocurrencia de un error tipo I se denomina nivel de significación y se denota por α. **Error tipo II:** se presenta al aceptar la hipótesis nula si esta no es verdadera.

Otra manera de observar los tipos de errores que se puede incurrir al probar una hipótesis estadística

Situaciones posibles al probar una hipótesis estadística

	Realidad	
Resultado de la prueba	H_0 es verdadera	H_0 es falsa
Aceptar H_0	Decisión correcta	Error tipo II
Rechazar H_0	Error tipo I	Decisión correcta

¿Cómo escoger el valor de alfa (α)?

Es de utilización general asignar un valor bajo para α, es decir, asignar un valor pequeño a la probabilidad de rechazar la hipótesis nula siendo verdadera. En la investigación en las ciencias de la salud los valores usuales de α son 0,05 y 0,01.

El valor del nivel de significación debe ser seleccionado previo a la realización de cualquier cálculo para decidir respecto a la hipótesis nula. El nivel de significación no puede cambiarse una vez que se han obtenido resultados adversos o contrarios a los esperados por el investigador.

PASO 3. Verificación de supuestos

Las conclusiones obtenidas son válidas siempre y cuando se cumplan ciertas condiciones o supuestos.

- Aquellas pruebas que requieran que la población o poblaciones involucradas se distribuyan según la forma de cierta distribución de probabilidad se denominan **PRUEBAS PARAMÉTRICAS**.
- Aquellas pruebas o contrastes que no exigen que la población o poblaciones involucradas se distribuyan según la forma de una distribución de probabilidad específica se denominan **PRUEBAS NO PARAMÉTRICAS**.

PASO 4. Regla de Decisión.

Se establecen reglas para tomar una decisión respecto a la hipótesis nula. Estas reglas involucran al nivel de significación y a la significación del valor del estadístico de prueba usado para el contraste. La significación del estadístico de prueba o p-valor, representa el valor de la probabilidad de obtener un valor más pequeño y/o más grande que el valor encontrado para el estadístico de prueba.

Tabla 12. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

PASO 5. Realizar los cálculos y tomar la decisión

Se calcula el valor de la significación del estadístico de prueba y se compara este con el nivel de significación. El resultado de la comparación permite la toma de una decisión respecto a la hipótesis nula.

PLANTEAMIENTO DE CONTRASTES DE HIPÓTESIS

CONTRASTE DE HIPÓTESIS DE UNA MEDIA (UNA MUESTRA)

Paso 1. Hipótesis estadísticas:

H_0: $\mu = \mu_0$

H_1: $\mu \neq \mu_0$

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Si se dispone de una variable cuantitativa y se desea conocer si el promedio de esa variable es estadísticamente igual o diferente a un valor dado, se utiliza la prueba **t de Student para una muestra.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

Con el fin de someter a contraste estas hipótesis, se hace clic en el menú **Analizar** y en el submenú **Comparar Medias**, se selecciona la opción **T para una muestra**.

CONTRASTE DE HIPÓTESIS DE DOS MUESTRAS RELACIONADAS O PAREADAS

Paso 1. Hipótesis estadísticas

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Si se dispone de dos variables cuantitativas y se desea conocer si los promedios de esas variables son estadísticamente iguales o diferentes, se utiliza la prueba **t de Student muestras pareadas.** Este es el caso típico de mediciones pretest y postest, o mediciones iniciales y finales.

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

Con el fin de someter a contraste estas hipótesis, se hace clic en el menú **Analizar** y en el submenú **Comparar Medias**, se selecciona la opción **t de Student para muestras relacionadas**.

CONTRASTE DE HIPÓTESIS DE DOS MUESTRAS INDEPENDIENTES

Paso 1. Hipótesis estadísticas

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Si el objetivo es conocer si existen diferencias estadísticamente significativas en el promedio de una variable dependiente cuantitativa con respecto a una variable cualitativa (dicotómica), se utiliza la prueba **t de Student para muestras independientes.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

Con el fin de someter a contraste estas hipótesis, se hace clic en el menú **Analizar** y en el submenú **Comparar Medias**, se selecciona la opción **T para muestras independientes.**

CONTRASTE DE HIPÓTESIS ANÁLISIS DE VARIANZA DE UN FACTOR

Paso 1. Hipótesis estadísticas

H_0: $\mu_1 = \mu_2 ... = \mu_k$

H_1: $\mu_1 \neq \mu_2 ... \neq \mu_k$

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Si el objetivo es conocer si existen diferencias estadísticamente significativas en el promedio cuando hay más de dos grupos o niveles en un solo factor, se utiliza la prueba **ANOVA de un factor.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

Con el fin de someter a contraste estas hipótesis, se hace clic en el menú **Analizar** y en el submenú **Comparar Medias**, se selecciona la opción **ANOVA de un factor.**

CONTRASTE DE HIPÓTESIS DE INDEPENDENCIA

Paso 1. Hipótesis estadísticas:

H_0: Existe relación estadísticamente significativa entre las variables involucradas.

H_1: No existe relación estadísticamente significativa entre las variables involucradas.

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Cuando se dispone de dos variables categóricas cualitativas nominales y ordinales, o variables que eran cuantitativas que para efecto de análisis se recodifican, y se desea determinar si existe relación estadísticamente significativa entre esas variables, la prueba estadística apropiada es **Chi Cuadrado.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor $\leq \alpha$	Rechazar la H_0
P-valor $> \alpha$	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

- Del menú **Analizar**, submenú **Estadísticos Descriptivos**, se selecciona la opción **Tablas de Contingencia.**
- Del listado se selecciona las variables, tanto para filas como para columnas.
- En el botón **Estadísticos**, se selecciona la opción **Chi Cuadrado**. Dependiendo de la escala de las variables también se dispone de otros coeficientes, los cuales serán utilizados de acuerdo a los objetivos de la investigación.
- En el botón Casillas, se pide mostrar frecuencias observadas, esperadas y porcentajes de fila, columna o total según los requerimientos del caso.

CONTRASTE DE HIPÓTESIS DE CORRELACIÓN DE PEARSON

Paso 1. Hipótesis estadísticas:

H_0: No Existe correlación entre las variables involucradas.

H_1: Existe correlación entre las variables involucradas.

Paso 2. Se establece el valor de α, puede ser $\alpha = 0{,}01$ o $\alpha = 0{,}05$

Paso 3. Verificación de supuestos

Cuando se desea determinar la significación estadística acerca del grado de asociación lineal entre dos variables expresadas en escala de intervalo o razón, la prueba estadística apropiada es el **Coeficiente de Correlación de Pearson.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor ≤ α	Rechazar la H_0
P-valor > α	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

- Del menú **Analizar**, submenú **Correlaciones**, se selecciona la opción Bivariadas.
- Del listado de variables se seleccionan las variables de interés, el **Coeficiente de Correlación de Pearson**, la dirección de la prueba de significación: bilateral o unilateral y la opción de presentar o no las correlaciones significativas.

CONTRASTE DE HIPÓTESIS DE CORRELACIÓN DE SPEARMAN

Paso 1. Hipótesis estadísticas

H_0: No existe correlación lineal entre las variables involucradas.

H_1: Existe correlación lineal entre las variables involucradas.

Paso 2. Se establece el valor de α, puede ser α = 0,01 o α = 0,05

Paso 3. Verificación de supuestos

Cuando se desea determinar la significación estadística acerca el grado de asociación lineal entre dos variables, una de ellas expresada en escala de intervalo o razón y la otra en escala ordinal, o ambas variables en escala ordinal, la prueba estadística apropiada es **Coeficiente de Correlación de Spearman.**

Paso 4. Regla de Decisión

Situación encontrada	Decisión
P-valor ≤ α	Rechazar la H_0
P-valor > α	No se rechaza la H_0

Paso 5. Cálculos con SPSS para Windows para tomar la decisión.

- Del menú **Analizar**, submenú **Correlaciones**, se selecciona la opción **Bivariadas**.

- Del listado de variables se seleccionan las variables de interés, el **Coeficiente de Correlación de Spearman**, la dirección de la prueba de significación: bilateral o unilateral y la opción de presentar o no las correlaciones significativas.

Referencias

Montgomery, D. y Runger, G. (2008). *Probabilidad y estadística aplicada a la ingeniería* (2a. Ed.). México: Limusa Wiley.

Ovalles, A. y Moret, C. (2001). *Manual de estadística ii*. Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

CAPÍTULO 10. MUESTREO

En una población donde se desea aplicar un análisis estadístico, cuando el muestreo cubre a todos los elementos, se realiza un **CENSO,** por lo tanto, es la medición de todos los elementos que constituyen la población.

¿Y SI NO PUEDO HACER UN CENSO? Se realiza un muestreo, esto ocurre cuando la Población es desconocida, inaccesible e inalcanzable, O también porque no existen los suficientes recursos económicos y humanos, la población suele destruirse y debido a que no hay suficiente tiempo.

El **MUESTREO:** es el proceso seguido para la extracción de una muestra. El objetivo del muestreo es considerar el mayor número de unidades con el menor costo posible.

ELECCIÓN DE LOS INDIVIDUOS A ESTUDIAR: al no existir registro de la información del fenómeno o experimento de interés, se debe recoger de la fuente original.

¿CÓMO HACER UNA ACERTADA ESCOGENCIA DE LOS ELEMENTOS A ESTUDIAR?
Se debe considerar las condiciones de **CALIDAD** y **CANTIDAD** de la muestra.

CANTIDAD: es el número óptimo y mínimo de elementos para el estudio. **Por ejemplo:** en un estudio de investigación epidemiológico la determinación de un tamaño adecuado de la muestra, ayuda a obtener resultados más confiables.

CALIDAD: la muestra debe reflejar las características de la población de la cual procede, solo diferenciándose en el número de unidades incluidas (**Nota**: este criterio es más importante que la cantidad y no lo contrario). **Ejemplo** de una muestra que no posee calidad: Si deseáramos estudiar las características del pueblo venezolano, y nos empeñáramos en estudiar solamente a los habitantes de Mérida, aunque estudiáramos a todos ellos, nuestra muestra no sería representativa de todo el país.

VENTAJAS DE UN MUESTREO: ahorrar tiempo, dinero y trabajo, además permite una mayor exactitud en el estudio.

DESVENTAJA DE UN MUESTREO: error de muestreo.

ERROR DE MUESTREO: es la diferencia entre el valor calculado en la muestra y el valor esperado de la población. Este tipo de erro puede invalidar el estudio. Condición: se deben utilizan idénticos métodos de estudio para la recolección de datos, de lo contrario serían errores por el método utilizado.

El **error de muestreo** es menos importante que errores debido al observador, por el método de observación, o por los individuos estudiados y este error puede medirse estadísticamente y puede disminuirse a voluntad con solo aumentar el tamaño de la muestra.

Ejemplo. Se tiene una población de 4 personas con Bs. 5, 3, 4 y 8, el capital total de esta población es Bs. 20, en promedio seria Bs. 5. Si no se conociera dicho promedio y para averiguarlo se toma una muestra de 2 personas, digamos las 2 primeras (Bs. 5 y Bs. 3), se concluiría a través de esa muestra que el promedio de la población es Bs. 4, cuando en realidad vimos que era 5. La diferencia entre el valor de la muestra y el valor de la población es el error de muestreo, en este caso Bs. 1

CONCEPTOS BÁSICOS DE MUESTREO

POBLACIÓN OBJETIVO: es la población total que se pretende estudiar.

POBLACIÓN DE ESTUDIO: es el grupo que en realidad se puede estudiar.

MARCO MUESTRAL: es un listado actualizado y revisado, de todos los elementos que constituyen la población que va a ser objeto de investigación. También puede ser un mapa o croquis con las unidades de selección plenamente identificadas.

UNIDAD DE MUESTREO: elementos o conjuntos no solapados de elementos los cuales cubren totalmente la población. Se utilizan para seleccionar la muestra. Pueden ser unidades simples (elementos, personas) o complejas (familias, organizaciones).

MUESTRA REPRESENTATIVA: es el grado en el cual la muestra reproduce las características de la población.

¿CUÁNDO UNA MUESTRA ES REPRESENTATIVA DE LA POBLACIÓN?

Requiere que la población sea homogénea y que todas las unidades de la población tengan la misma probabilidad de ser seleccionada (Martínez, 2008). Es fundamental considerar la cantidad y la calidad de la muestra (técnica de muestreo). En otras palabras, las tres condiciones son: homogeneidad, aleatoriedad y técnica de muestreo bien definida (cantidad y calidad).

¿CÓMO ELEGIR EL MUESTREO APROPIADO?

Se recomiendan **muestreos no probabilísticos**, por limitaciones de recursos, tiempo, dinero y trabajo, se debe estudiar un número de individuos menor que el deseable y entonces la opinión del experto se hace conveniente. Los **muestreos probabilísticos,** en la literatura se menciona que deberían utilizarse siempre que sean posibles de realizar, sin embargo veamos algunas excepciones:

Ejemplo. Si se desea ensayar una nueva droga y solamente se tienen 5 o 6 dosis, en lugar de escoger los individuos al azar pueden seleccionarse casos graves, ya que si se mejoran estos casos, será válido para pacientes con menor o sin gravedad.

Ejemplo. En ocasiones no se puede obtener una lista completa de la población que se va a estudiar, siendo por lo tanto imposible aplicar el azar. En dicho caso, la selección de los individuos que se estudian envuelve un proceso de opinión.

Ejemplo. En ocasiones el principal interés está en localizar individuos con determinadas características en una población muy numerosa, digamos los enfermos tuberculosos de una colectividad. En tal caso, es preferible concentrarnos en el estudio de aquellos grupos en los cuales la experiencia señala que hay más probabilidad de encontrar los individuos buscados.

MUESTREO NO PROBABILÍSTICO: es un procedimiento por medio del cual las unidades muestrales no se seleccionan al azar, sino que son elegidas por el responsable de realizar el muestreo. La selección de la muestra se basa en el criterio del investigador. El costo de dichos muestreos es menor comparado con un muestreo probabilístico.

Este tipo de muestreo estriba en la posibilidad de que un individuo sea incluido en la muestra desconocida, siendo imposible medir la exactitud de los resultados obtenidos, porque no se puede medir el error o nivel de confianza, porque no se pueden incluir ecuaciones de probabilidad.

DESVENTAJAS DEL MUESTREO NO PROBABILÍSTICO

- Incapacidad de juzgar la precisión de la muestra.
- Mecanismo poco objetivo de apreciación.
- No ofrece representatividad.
- No se puede medir la exactitud de los resultados.

TIPOS DE MUESTREOS NO PROBABÍLISTICO

A JUICIO, INTENCIONAL U OPINÁTICO: los elementos son seleccionados a juicio o en opinión del investigador.

POR CONVENIENCIA: se eligen las unidades de muestreo que se encuentran a mayor alcance del investigador.

VOLUNTARIADO: el individuo voluntariamente suministra información sin ser seleccionado.

MUESTREO PROBABILÍSTICO: es un procedimiento en el cual cada individuo de la población, tiene probabilidad perfectamente conocida, por lo tanto, cada individuo es seleccionado considerando aleatoriedad. No es necesario que los individuos cumplan con el principio de equiprobabilidad, basta con que tenga cualquier posibilidad diferente de cero de formar parte de la muestra y que esa probabilidad sea conocida. Parten del principio de que todas las posibles muestras de tamaño n tienen la misma probabilidad de ser elegidas. Estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables.

CONDICIONES DE UN MUESTREO PROBABILÍSTICO

- La probabilidad de elegir cada individuo sea perfectamente conocida, de lo contrario, NO se podrán calcular los errores al momento de la selección.
- Es fundamental que los individuos se elijan al azar, se puede usar, por ejemplo: la tabla de números aleatorios, el método de la lotería u otro método.

MUESTREO ALEATORIO SIMPLE: procedimiento donde todos y cada uno de los elementos de la población tienen la misma probabilidad de ser seleccionados en la muestra y esta probabilidad es conocida. Este tipo de muestreo es más recomendable, pero resulta mucho más difícil de llevarse a cabo y por lo tanto, es más costoso. Para seleccionar una muestra de este tipo se requiere tener en forma de lista todos los elementos que integran la población investigada y utilizar algún instrumento, yal como las tablas de números aleatorios.

Ejemplo. Suponga que estamos investigando sobre el porcentaje de estudiantes que fuman según el sexo de una población de 20 estudiantes de la Universidad de Los Andes

Base de datos de la población:

Nombre Alumno	¿Fuma?
Juan	SI
Alicia	NO
Pedro	NO
Marcos	NO
Alberto	SI
Jorge	SI
José	NO
Carlos	NO
Miguel	NO
Victoria	SI

Nombre Alumno	¿Fuma?
María	NO
Fernanda	NO
Julio	SI
Rosa	NO
Fabián	NO
Ana	NO
Laura	NO
Enrique	NO
Carmen	SI
Marcelo	SI

Paso 1. Asignamos un número a cada estudiante del 1 al 20:

Número	Nombre Alumno	¿Fuma?
1	Juan	SI
2	Alicia	NO
3	Pedro	NO
4	Marcos	NO
5	Alberto	SI
6	Jorge	SI
7	José	NO
8	Carlos	NO
9	Miguel	NO
10	Victoria	SI

Número	Nombre Alumno	¿Fuma?
11	María	NO
12	Fernanda	NO
13	Julio	SI
14	Rosa	NO
15	Fabián	NO
16	Ana	NO
17	Laura	NO
18	Enrique	NO
19	Carmen	SI
20	Marcelo	SI

Paso 2. Buscamos en la tabla de números aleatorios 4 números, de dos dígitos, entre el 1 y el 20, sin repetir:

Tabla de números aleatorios:

columna / fila	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207
2	22368	46573	25595	85393	30995	89198	37982	53402	93965	34095
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672

Los números seleccionados son: 10, 1, 11 y 20. Por lo tanto, la muestra está compuesta por Victoria, Juan y Marcelo que fuman y María que no fuma.

Paso 3. Conclusiones

- El 75% de los estudiantes de la Universidad de Los Andes fuman.
- El 50% de los estudiantes de la Universidad de Los Andes son hombres fumadores.
- El 25% de los estudiantes de la Universidad de Los Andes son mujeres fumadoras.

¿Qué falló en el muestreo para que se dieran resultados no extrapolables a la población?
Fundamentalmente falló el tamaño de la muestra.

¿CÓMO CALCULAR EL TAMAÑO DE LA MUESTRA?
Fundamentalmente falló el tamaño de la población y muestra, pues ante poblaciones pequeñas se puede hacer un censo.

VARIANZA (σ^2): Correspondiente al grado de variabilidad que presentan las unidades de la población. Mientras más grande sea la varianza, mayor será el tamaño de la muestra. El valor de la Varianza se debe conocer, de lo contrario se debe estimar a través de una investigación preliminar. En el caso de la Varianza de una proporción, se toma P=0,5, con lo cual se obtiene el máximo valor posible de n (Martínez, 2008).

NIVEL DE CONFIANZA: Tiene relación directa con el tamaño de la muestra, por lo tanto se dirá que a mayor nivel de confianza más grande debe ser el tamaño de la muestra. Los valores de la

Distribución Normal Estandarizada (Z) se obtienen mediante el uso de tablas. El nivel es fijado por el investigador de acuerdo con su experiencia.

PRECISIÓN DE LA ESTIMACIÓN: Corresponde al margen de error que el investigador fija de acuerdo con el conocimiento que tenga acerca del parámetro que piensa estimar. Se le conoce como error de muestreo (E).

Ejemplo. Supongamos que se quiere obtener una muestra para la población de estudiantes del Ejemplo 1, con los siguientes datos, con un Nivel de Confianza del 95% que en la tabla de Z es 1,96 y se estima que P=0,1, por lo tanto Q=0,9 y se asume un E=0,02

$$n = \frac{Z^2 NPQ}{NE^2 + Z^2 PQ}$$

$$n = \frac{1{,}96^2 \times 20 \times 0{,}1 \times 0{,}9}{(20) \times (0{,}02)^2 + 1{,}96^2 \times (0{,}1) \times (0{,}9)} = 19{,}54 \approx 20$$

Por ello se recomienda hacer un censo ante poblaciones pequeñas.

Ejemplo. Poblaciones infinitas.

Un médico desea investigar sobre los accidentes de motos, para ello quiere tomar una muestra con un nivel de confianza del 99% y que no exceda un error del 2% ¿Qué tamaño de muestra tendrá que tomar si estima que la proporción del error es del 8%?

Solución:

Nivel de Confianza del 99% que en la tabla de Z es 2,58

P=0,08, por lo tanto Q=0,92.

E=0,02

$$n = \frac{Z^2 PQ}{E^2} = \frac{2{,}58^2 \times (0{,}08) \times (0{,}92)}{0{,}02^2} = 1224{,}78 \approx 1225$$

Ejemplo. Error muestral.

De un conjunto de gorros descartables se tomaron una muestra de 200, se encontró que 9 de ellos eran defectuosos. Con una confianza del 95%, calcular el error de la muestra.

Solución:

$$P = \frac{9}{200} = 0{,}045$$

$$E = Z\sqrt{\frac{PQ}{n}} = 1{,}96\sqrt{\frac{(0{,}045) \times (0{,}955)}{200}} = 0{,}0287$$

Expresado en porcentaje el error muestral es del 2,87%

Ejemplo. Tamaño de poblaciones finitas.

El INTTT desea tomar una muestra para estimar la proporción de conductores con experiencia de 1 año o menos, que puedan clasificarse como conductores descuidados ¿De qué tamaño es la muestra si se considera 10 mil conductores a investigar, utilizando un nivel de confianza del 95% y un error muestral del 2%? Se espera observar que aproximadamente ¼ de los conductores sean descuidados.

Solución:

$$n = \frac{Z^2NPQ}{NE^2 + Z^2PQ}$$

$$n = \frac{1{,}96^2 \times 10000 \times 0{,}25 \times 0{,}75}{(10000) \times (0{,}02)^2 + 1{,}96^2 \times (0{,}25) \times (0{,}75)} = 1526$$

MUESTREO SISTEMÁTICO: La selección de unidades se halla a través intervalos regulares en un orden sistemático. La lista de elementos debe estar realizada al azar. El punto de partida debe ser al azar. **CUIDADO:** Si en la lista existen periodicidades, se obtendría una muestra sesgada.

Ejemplo. En la Facultad de Medicina de la Universidad de Los Andes, se desea elegir una muestra sistemática de 30 estudiantes a partir de una población de 120 estudiantes que poseen enfermedades respiratorias.

Solución

Paso 1. Se enumeran los estudiantes.

Paso 2. Se calcula la constante (k) entre cada intervalo, es decir:

$$k = \frac{N}{n} = \frac{120}{30} = 4$$

Paso 3. Se sortea un número del 1 al 4, a partir del número obtenido al azar se le suma la constante hasta conseguir la cantidad de la muestra.

Supongamos que sea 2, entonces la muestra queda conformada por los siguientes números:

$$\{2, 6,10,14,18,22,26,30,34,38,42,46,50,54,58,62,66,70,74,78,82,86,90,94,98,102,106,110,114,118\}$$

Paso 4. Enuncie las conclusiones de la población a partir de la muestra.

MUESTREO ESTRATIFICADO: En este procedimiento la población se divide en estratos y luego en cada uno de los estratos se escogen al azar los individuos que compondrán la muestra, haciendo una asignación homogénea, es decir, proporcional por cada estrato de acuerdo a la composición de la población. **DIFICULTAD:** exige un conocimiento muy detallado de la población.

VENTAJAS DEL MUESTREO ESTRATIFICADO

- Se obtiene información separada de cada uno de los estratos.
- Se evita el riesgo de que determinada muestra quede inadecuadamente representada.

Ejemplo. Es conveniente el muestreo estratificado, en el caso de los días de hospitalización de los Servicios de Ginecología y Obstetricia, Pediatría, Cirugía y Medicina general, son diferentes unos de otros. En este caso se hace una muestra separada de cada uno de los 4 servicios y luego se combinan los resultados.

Ejemplo. Se está investigando sobre el porcentaje de estudiantes (N=120) fumadores en la Universidad de Los Andes, para ello se selecciona una muestra estratificada por sexo (masculino 30 estudiantes y femenino 90 estudiantes) con un nivel de confianza del 95%, Error del 3% y P=0,15.

Utilice la tabla de números aleatorios en cada estrato, comenzando en la fila 1, columna 1 y continúe seleccionando hacia la derecha. Indique los pasos para elegir la muestra.

Solución. Procedimiento

Paso 1. Para elegir una muestra estratificada, se dividen los estratos y se le asignan un número de identificación.

Paso 2. Se determina la proporción por sexo, es decir:

N = 120 estudiantes. Estrato de hombres: 30. Estrato de mujeres: 90

Proporción de hombres = 30/120 = 0,25

Proporción de mujeres = 90/120 = 0,75

Paso 3. Halle el tamaño de la muestra

$$n = \frac{1{,}96^2 \times 120 \times 0{,}15 \times 0{,}85}{(120) \times (0{,}03)^2 + 1{,}96^2 \times (0{,}15) \times (0{,}85)} = 98{,}32 \approx 99$$

Paso 4. Determine la cantidad por estrato:

Muestra de hombres = 0,25 x 99 = 24,75 ≈ 25

Muestra de mujeres = 0,75 x 99 = 74,25 ≈ 74

Paso 5. Utilizando la tabla de números aleatorios, se elige una muestra aleatoria simple de tamaño n=25 para los hombres, buscando números del 1 al 25. Se parte de la fila 1, columna 1.

Paso 6. Empleando la tabla de números aleatorios, se elige una muestra aleatoria simple de tamaño n=75 para las mujeres, buscando números del 1 al 75. Se parte de la fila 1, columna 1.

Paso 8. Enuncie las conclusiones de la población a partir de la muestra.

MUESTREOS POR CONGLOMERADOS. En lugar de escogerse a las unidades que van a estudiarse, se selecciona un grupo o conglomerado de elementos. No es necesario conocer a todos los individuos a estudiar, basta con enumerar a los grupos o conglomerados a investigar.

Procedimiento

- Dividir la población en conglomerados.
- Seleccionar al azar el número de conglomerados según la necesidad y capacidad del investigador.
- Se calcula el tamaño de muestra.
- Tomar una muestra aleatoria simple o un muestreo sistemático de uno de los elementos de cada conglomerado seleccionado.

Ejemplo. Se desea efectuar una encuesta sobre las políticas de salud a las familias del municipio Libertador (Mérida-Venezuela).

Procedimiento

1. Se divide el municipio en distritos, por ejemplo en 10 distritos.
2. De la división efectuada se toman al azar los distritos 4, 5, 7 y 10.
3. Se calcula el tamaño de muestra.
3. Se realiza un muestreo sistemático, para ejecutar la recolección de datos en las familias por vivienda de los distritos del municipio Libertador seleccionados.

MUESTREOS COMBINADOS: Es la forma de muestreo que resulta de combinar en varias etapas, dos o más de los métodos antes descritos. **Ejemplo.** Para un estudio sobre Enfermedades de Transmisión Sexual en un municipio, se selecciona al azar 20 de los consultorios Médicos del área urbana y 20 del área rural. Posteriormente, se toman los registros, de día por medio, durante 2 semanas.

Referencias

Camel, F. (1991). Estadística Médica y Planificación de la Salud. Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Martínez, C. (2008). *Estadística y muestreo* (12da. Ed.). Bogotá: ECOE Ediciones.

REFERENCIAS CONSULTADAS

Armas, J. (1988). *Estadística sencilla descriptiva.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Camel, F. (1991). *Estadística Médica y Planificación de la Salud.* Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Chipia, J. y Lara, C. (2008). *Módulo para la enseñanza–aprendizaje de la estadística en el bachillerato mediante la solución de situaciones problemas.* Mérida: Tesis de Grado mención Publicación en la Universidad de Los Andes.

Chipia, J. (2009). *Propuesta para la enseñanza de la estadística en primer año de secundaria mediante resolución de problemas.* Revista Voces: Tecnología y pensamiento, 4(1-2), 79-96.

Chipia, J. (2012). *Gráficas de caja* [blog]. Bioestadística y Matemática Básica #ULA. Disponible: http://bioestadisticaula.blogspot.com/2012/09/graficas-de-caja.html [Consulta: 2015 Marzo 30].

Chipia, J. (2012). *Gráfico de puntos y pictogramas* [blog]. Bioestadística y Matemática Básica #ULA. Disponible: http://bioestadisticaula.blogspot.com/2012/07/grafico-de-puntos-y-pictogramas.html [Consulta: 2015 Marzo 30].

Chipia, J., Cadenas, R. y Lara, C. (2012) *Propuesta para la enseñanza de organización de datos para variables cualitativas.* Revista EDUCERE, 16 (53), 185-196.

Chipia, J. (2013). *Efectividad de un Programa de Enseñanza/Aprendizaje sobre Estadística Descriptiva utilizado Calc de Open.* Memoria en extenso (N°45) del VII CIBEM, 16-20 de septiembre de 2013.

Chipia, J. (2014). *Efectividad de un programa de enseñanza/aprendizaje sobre estadística descriptiva.* Mérida: Tesis de Maestría mención Honorífica en la Universidad de Los Andes.

Daniel, W. (2010). *Bioestadística: Base para el análisis de las ciencias de la salud* (4a. Ed.). México: Limusa Wiley.

Glass, G. y Stanley, J. (1986). *Métodos estadísticos aplicados a las ciencias sociales*. México: Prentice Hall.

Hays, W. L. (1973). *Statistics for the social sciences* (2da. Ed.). New York: Holt, Rinehart and Wilson.

Hernández, R. (2011). *Variabilidad absoluta y relativa en distribuciones de frecuencias.* Mérida: Consejo de Estudios de Postgrado de la Universidad de Los Andes.

Hernández, S. (2005). *Historia de la estadística.* Revista de divulgación científica y tecnológica de la Universidad Veracruzana [En línea], 18, (8). Disponible: http://www.uv.mx/cienciahombre/revistae/vol18num2/articulos/historia/index.htm [Consulta: 2012 Abril 20].

Martínez, C. (2008). *Estadística y muestreo* (12da. Ed.). Bogotá: ECOE Ediciones.

Montgomery, D. y Runger, G. (2008). *Probabilidad y estadística aplicada a la ingeniería* (2a. Ed.). México: Limusa Wiley.

Ovalles, A. y Moret, C. (2001). *Manual de estadística ii*. Mérida: Consejo de Publicaciones de la Universidad de Los Andes.

Anexo A. Nacimientos vivos registrados, según entidad federal de residencia habitual de la madre, 2000-2010

Entidad Federal	Año de registro										
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Total	544.416	529.5	492.678	555.614	637.799	665.997	646.225	615.371	581.480	593.845	591.303
Distrito Capital	37.169	31.5	29.229	38.886	42.048	42.476	44.787	43.710	38.394	38.498	37.896
Amazonas	3.020	4.55	4.421	3.454	3.478	4.578	5.338	5.341	4.728	4.933	4.868
Anzoátegui	30.799	28.7	28.061	31.712	31.759	38.380	35.619	34.873	33.383	32.567	32.259
Apure	12.560	14.4	11.766	14.517	16.165	18.580	15.916	14.446	14.456	14.549	16.161
Aragua	26.962	28.1	27.574	27.112	29.576	29.846	28.765	27.588	26.900	30.999	30.616
Barinas	17.173	15.9	16.379	16.782	22.289	20.684	19.425	18.621	18.412	18.643	19.431
Bolívar	33.993	30.2	27.079	31.740	38.565	40.223	39.171	41.155	36.190	38.346	39.591
Carabobo	39.704	43.2	35.706	43.778	48.731	47.286	49.152	48.119	46.383	45.631	46.673
Cojedes	7.065	6.47	6.201	5.920	7.003	4.907	4.031	3.863	3.830	4.372	6.838
Delta Amacuro	4.290	3.41	3.466	6.013	6.213	7.405	5.796	6.801	4.440	5.519	3.901
Falcón	18.335	17.3	16.457	17.594	20.670	20.798	18.411	19.276	18.345	17.362	18.032
Guárico	15.260	14.7	15.439	16.438	20.243	19.720	17.848	17.821	19.310	17.224	17.266
Lara	33.413	33.6	26.997	43.835	38.620	37.051	35.383	28.625	30.208	34.671	31.479
Mérida	19.672	15.5	15.391	15.849	17.622	17.976	16.137	16.672	16.826	16.968	16.940
Miranda	49.936	47.1	43.606	49.091	45.149	52.470	60.899	56.985	51.050	58.275	52.090
Monagas	17.630	18.8	15.958	18.429	23.090	22.645	22.061	24.668	19.516	21.047	23.164
Nueva Esparta	8.356	7.89	8.300	8.210	9.555	11.171	9.471	9.955	9.757	9.824	9.852
Portuguesa	18.690	15.0	18.729	18.650	22.083	25.508	20.857	18.900	18.217	16.488	20.087
Sucre	22.974	18.2	17.322	18.816	20.489	23.982	19.791	19.872	18.488	20.139	20.747
Táchira	23.101	20.8	16.352	17.708	21.118	20.496	21.050	16.575	19.994	20.205	22.013
Trujillo	14.485	14.1	13.409	11.799	12.330	19.485	14.748	14.681	14.069	14.658	14.528
Vargas	6.676	7.21	6.852	7.003	6.132	7.324	7.588	6.004	6.198	6.030	6.839
Yaracuy	12.214	12.1	11.091	12.419	13.334	15.501	14.111	11.759	11.367	11.112	11.831
Zulia	70.939	79.9	76.893	79.859	121.53	117.50	119.87	109.06	101.01	95.785	88.201

Fuente: Oficinas de registro civil municipal y unidades de registro civil parroquial y en establecimientos de salud, INE 2011. Tomado de http://www.ine.gov.ve/documentos/Demografia/EstadisticasVitales/html/NatEntFedResMad.html

Anexo B. Base de datos

N°	*Edad (años)*	*Estatura (m)*	*Masa (Kg)*	*Nivel Educativo*	*¿Fuma?*	*Sexo*	*Hemoglobina (g/ml)*
1	5	1,02	20	Inicial	No	Femenino	10,5
2	18	1,56	55	Media General	Si	Masculino	11,5
3	17	1,64	66	Media General	No	Femenino	11,3
4	15	1,65	67	Media General	No	Femenino	10,9
5	14	1,80	76	Media General	No	Femenino	10,8
6	10	1,29	39	Primaria	No	Femenino	11,0
7	9	1,20	38	Primaria	No	Masculino	11,5
8	12	1,55	54	Media General	No	Femenino	12,0
9	45	1,67	67	Media General	Si	Femenino	11,8
10	56	1,63	60	Universitario	No	Masculino	12,8
11	66	1,58	70	Media General	No	Masculino	13,0
12	46	1,61	71	Universitario	No	Femenino	13,0
13	88	1,50	55	Primaria	No	Femenino	12,0
14	12	1,41	54	Media General	No	Femenino	12,1
15	76	1,51	70	Media General	Si	Femenino	9,6
16	55	1,67	60	Media General	Si	Femenino	9,9
17	43	1,80	40	Universitario	No	Masculino	10,8
18	26	1,57	55	Primaria	No	Masculino	10,7
19	35	1,78	79	Universitario	No	Femenino	11,8
20	38	1,45	49	Universitario	Si	Masculino	10,5
21	70	1,74	76	Primaria	No	Femenino	10,9
22	22	1,59	63	Primaria	Si	Femenino	11,5
23	2	0,89	12	Maternal	No	Femenino	12,3
24	54	1,66	78	Universitario	Si	Femenino	12,9
25	34	1,78	89	Primaria	No	Femenino	12,1
26	35	1,69	44	Primaria	Si	Masculino	11,9
27	67	1,73	66	Media General	No	Femenino	11,3
28	80	1,65	77	Universitario	No	Masculino	10,8
29	12	1,59	62	Media General	No	Femenino	11,1
30	11	1,66	63	Media General	No	Masculino	11,2
31	34	1,54	59	Media General	No	Masculino	12,5
32	66	1,67	73	Media General	Si	Femenino	12,0
33	90	1,49	48	Media General	Si	Femenino	10,9
34	3	0,99	15	Inicial	No	Femenino	10,0
35	2	0,87	13	Inicial	No	Femenino	10,1
36	1	0,63	10	Media General	No	Masculino	10,0
37	2	0,81	13	Inicial	No	Masculino	10,2
38	5	1,01	15	Inicial	No	Masculino	10,3
39	4	1,03	18	Inicial	No	Masculino	10,6
40	7	1,23	30	Primaria	No	Masculino	11,3
41	34	1,67	66	Media General	No	Masculino	12,7
42	45	1,78	77	Media General	Si	Femenino	14,0
43	55	1,72	70	Ninguno	Si	Femenino	13,5
44	43	1,59	60	Ninguno	Si	Masculino	13,6
45	55	1,55	77	Ninguno	Si	Masculino	13,8
46	66	1,54	49	Ninguno	Si	Femenino	12,9
47	70	1,63	66	Primaria	Si	Masculino	13,2
48	25	1,71	70	Ninguno	Si	Femenino	13,3
49	20	1,72	85	Primaria	Si	Masculino	12,4
50	21	1,49	55	Ninguno	No	Masculino	8,9

51	21	1,51	56	Media General	No	Femenino	10,5
52	23	1,54	58	Primaria	No	Masculino	9,6
53	24	1,50	60	Media General	No	Masculino	11,7
54	55	1,61	62	Media General	Si	Masculino	11,5
55	46	1,59	58	Media General	No	Femenino	9,0
56	49	1,55	57	Media General	Si	Masculino	9,9
57	54	1,54	64	Media General	No	Masculino	10,7
58	46	1,63	67	Media General	Si	Femenino	10,8
59	71	1,71	73	Media General	No	Femenino	11,3
60	33	1,72	75	Media General	Si	Femenino	11,9
61	44	1,49	51	Media General	No	Femenino	12,3
62	35	1,51	51	Media General	No	Femenino	12,6
63	27	1,54	55	Primaria	No	Femenino	12,7
64	26	1,50	55	Primaria	No	Masculino	9,6
65	21	1,61	55	Media General	No	Masculino	9,8
66	23	1,78	77	Media General	Si	Masculino	9,9
67	45	1,79	77	Primaria	Si	Masculino	10,6
68	47	1,69	70	Media General	No	Masculino	11,2
69	49	1,90	88	Media General	Si	Femenino	13,1
70	50	1,87	89	Media General	No	Femenino	12,7
71	55	1,50	55	Primaria	Si	Masculino	14,7
72	58	1,57	59	Primaria	No	Masculino	13,6
73	60	1,61	55	Universitario	No	Femenino	13,3
74	66	1,63	60	Media General	No	Masculino	10,1
75	70	1,76	66	Universitario	No	Femenino	11,5
76	4	1,77	77	Primaria	Si	Femenino	12,5
77	5	1,80	25	Media General	Si	Masculino	13,6
78	12	1,89	40	Inicial	No	Masculino	13,8
79	16	1,99	45	Primaria	No	Femenino	14,1
80	22	1,55	48	Primaria	Si	Femenino	14,6
81	40	1,72	75	Media General	Si	Femenino	11,9
82	30	1,49	51	Media General	No	Femenino	12,3
83	20	1,51	51	Media General	No	Femenino	12,6
84	7	1,04	55	Primaria	No	Femenino	12,7
85	37	1,50	55	Primaria	No	Masculino	9,6
86	28	1,61	55	Media General	No	Masculino	9,8
87	17	1,78	57	Media General	Si	Masculino	9,9
88	90	1,69	70	Primaria	Si	Masculino	10,6
89	89	1,59	70	Media General	No	Masculino	11,2
90	77	1,80	88	Media General	Si	Femenino	13,1
91	50	1,67	89	Media General	No	Femenino	12,7
92	60	1,50	55	Primaria	Si	Masculino	14,7
93	5	1,01	38	Inicial	No	Masculino	13,6
94	5	1,02	45	Inicial	No	Femenino	13,3
95	4	0,99	34	Inicial	No	Masculino	10,1
96	70	1,76	66	Universitario	No	Femenino	11,5
97	30	1,77	77	Primaria	Si	Femenino	12,5
98	29	1,80	25	Media General	Si	Masculino	13,6
99	18	1,89	40	Inicial	No	Masculino	13,8
100	16	1,99	45	Primaria	No	Femenino	14,1

Nota: datos hipotéticos de pacientes atendidos en el IAHULA, Mérida, Venezuela.

Anexo C. Números Aleatorios

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02
08	42	26	89	53	19	64	50	93	03	23	20	90	25	00	15	95	33	47	64
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	08	77
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85
31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39
85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47
63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09
73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44
98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33
11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01
83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10
88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93
99	59	46	73	48	87	51	76	49	69	91	82	60	89	28	93	78	56	13	68
65	48	11	76	74	17	46	85	09	50	58	04	77	69	74	73	03	95	71	86
80	12	43	56	35	17	72	70	80	15	45	31	82	23	74	21	11	57	82	53
74	35	09	98	17	77	40	27	72	14	43	23	60	02	10	45	52	16	42	37
69	91	62	68	03	66	25	22	91	48	36	93	68	72	03	76	62	11	39	90
09	89	32	05	05	14	22	56	85	14	46	42	75	67	88	96	29	77	88	22
91	49	91	45	23	68	47	92	76	86	46	16	28	35	54	94	75	08	99	23
80	33	69	45	98	26	94	03	68	58	70	29	73	41	35	53	14	03	33	40
44	10	48	19	49	85	15	74	79	54	32	97	92	65	75	57	60	04	08	81
12	55	07	37	42	11	10	00	20	40	12	86	07	46	97	96	64	48	94	39
63	60	64	93	29	16	50	53	44	84	40	21	95	25	63	43	65	17	70	82
61	19	69	04	46	26	45	74	77	74	51	92	43	37	29	65	39	45	95	93
15	47	44	52	66	95	27	07	99	53	59	36	78	38	48	82	39	61	01	18
94	55	72	85	73	67	89	75	43	87	54	62	24	44	31	91	19	04	25	92
42	48	11	62	13	97	34	40	87	21	16	86	84	87	67	03	07	11	20	59
23	52	37	83	17	73	20	88	98	37	68	93	59	14	16	26	25	22	96	63
04	49	35	24	94	75	24	63	38	24	45	86	25	10	26	61	96	27	93	36
00	54	99	76	54	64	05	18	81	59	96	11	96	38	96	54	69	28	23	91
35	96	31	53	07	26	89	90	93	54	33	35	13	54	62	77	97	45	00	24
59	80	80	83	91	43	42	72	68	42	83	60	94	97	00	13	02	12	48	92
46	05	88	52	36	01	39	09	22	86	77	28	14	40	77	93	91	08	36	47
32	17	90	05	97	87	37	92	52	41	05	56	70	70	07	86	74	31	71	57
69	23	46	14	06	20	11	74	52	04	15	95	66	00	00	18	74	39	24	23
19	56	54	14	30	01	75	87	53	79	40	41	92	15	85	66	67	43	68	06
45	15	51	49	38	19	47	60	72	46	43	66	79	45	43	59	04	79	00	33
94	86	43	19	94	36	16	81	08	51	34	88	88	15	53	01	54	03	54	56
59	58	00	64	78	75	56	97	88	00	88	83	55	44	86	23	76	80	61	56
38	50	80	73	41	23	79	34	87	63	90	82	29	70	22	17	71	90	42	07
30	69	27	06	68	94	68	81	61	27	56	19	68	00	91	82	06	76	34	00
65	44	39	56	59	18	28	82	74	37	49	63	22	40	41	08	33	76	56	76
27	26	75	02	64	13	19	27	22	91	07	47	74	46	06	17	98	54	89	11

Anexo D. Fórmulas de Estadística Descriptiva

- Rango Empírico (Re): Re= (Valor máximo – Valor mínimo) + Unidad
- Número de clases (Nc): $Nc = \sqrt{n}$
- Amplitud del Intervalo (a_i): $a_i = \frac{Re}{Nc}$
- Límite Aparente Inferior (LAI): LAI = Valor mínimo
- Límite Aparente Superior (LAS): LAS = (LAI + a_i) – Unidad
- Límite Real Inferior (LRI): LRI = LAI – (Unidad/2)
- Límite Real Superior (LRS): LRS = LAS + (Unidad/2)
- Percentiles (P_x): $P_{x=}LRI + \left(\frac{\left(\frac{n.x}{100}\right) - Ni(a)}{n_i}\right).a_i$
- Media Aritmética ($\bar{x}$): $\bar{x} = \frac{\sum Xm.n_i}{n}$
- Mediana (Md): $Md = LRI + \left(\frac{\left(\frac{n}{2}\right) - N_i(a)}{n_i}\right).a_i$
- Varianza (S^2): $S^2 = \frac{\sum(Xm - \bar{x})^2.n_i}{n}$
- Desviación Típica o Estándar (S): $S = \sqrt{S^2}$
- Coeficiente de Variación (CV):

Proporcional: $CV = \frac{S}{\bar{x}}$

Porcentual: $CV = \frac{S}{\bar{x}}.100$

- Error Típico de la Media (ETM): $ETM = \frac{S}{\sqrt{n}}$

Anexo E. Tabla de la Distribución Normal Estándar (Negativa)

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-4	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
-3,9	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,8	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
-3,7	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002
-3,6	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
-3,5	0,0002	0,0002	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003
-3,4	0,0003	0,0003	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0005	0,0005
-3,3	0,0005	0,0005	0,0005	0,0005	0,0006	0,0006	0,0006	0,0006	0,0006	0,0007
-3,2	0,0007	0,0007	0,0007	0,0008	0,0008	0,0008	0,0008	0,0009	0,0009	0,0009
-3,1	0,0010	0,0010	0,0010	0,0011	0,0011	0,0011	0,0012	0,0012	0,0013	0,0013
-3	0,0013	0,0014	0,0014	0,0015	0,0015	0,0016	0,0016	0,0017	0,0018	0,0018
-2,9	0,0019	0,0019	0,0020	0,0021	0,0021	0,0022	0,0023	0,0023	0,0024	0,0025
-2,8	0,0026	0,0026	0,0027	0,0028	0,0029	0,0030	0,0031	0,0032	0,0033	0,0034
-2,7	0,0035	0,0036	0,0037	0,0038	0,0039	0,0040	0,0041	0,0043	0,0044	0,0045
-2,6	0,0047	0,0048	0,0049	0,0051	0,0052	0,0054	0,0055	0,0057	0,0059	0,0060
-2,5	0,0062	0,0064	0,0066	0,0068	0,0069	0,0071	0,0073	0,0075	0,0078	0,0080
-2,4	0,0082	0,0084	0,0087	0,0089	0,0091	0,0094	0,0096	0,0099	0,0102	0,0104
-2,3	0,0107	0,0110	0,0113	0,0116	0,0119	0,0122	0,0125	0,0129	0,0132	0,0136
-2,2	0,0139	0,0143	0,0146	0,0150	0,0154	0,0158	0,0162	0,0166	0,0170	0,0174
-2,1	0,0179	0,0183	0,0188	0,0192	0,0197	0,0202	0,0207	0,0212	0,0217	0,0222
-2	0,0228	0,0233	0,0239	0,0244	0,0250	0,0256	0,0262	0,0268	0,0274	0,0281
-1,9	0,0287	0,0294	0,0301	0,0307	0,0314	0,0322	0,0329	0,0336	0,0344	0,0351
-1,8	0,0359	0,0367	0,0375	0,0384	0,0392	0,0401	0,0409	0,0418	0,0427	0,0436
-1,7	0,0446	0,0455	0,0465	0,0475	0,0485	0,0495	0,0505	0,0516	0,0526	0,0537
-1,6	0,0548	0,0559	0,0571	0,0582	0,0594	0,0606	0,0618	0,0630	0,0643	0,0655
-1,5	0,0668	0,0681	0,0694	0,0708	0,0721	0,0735	0,0749	0,0764	0,0778	0,0793
-1,4	0,0808	0,0823	0,0838	0,0853	0,0869	0,0885	0,0901	0,0918	0,0934	0,0951
-1,3	0,0968	0,0985	0,1003	0,1020	0,1038	0,1056	0,1075	0,1093	0,1112	0,1131
-1,2	0,1151	0,1170	0,1190	0,1210	0,1230	0,1251	0,1271	0,1292	0,1314	0,1335
-1,1	0,1357	0,1379	0,1401	0,1423	0,1446	0,1469	0,1492	0,1515	0,1539	0,1562
-1	0,1587	0,1611	0,1635	0,1660	0,1685	0,1711	0,1736	0,1762	0,1788	0,1814
-0,9	0,1841	0,1867	0,1894	0,1922	0,1949	0,1977	0,2005	0,2033	0,2061	0,2090
-0,8	0,2119	0,2148	0,2177	0,2206	0,2236	0,2266	0,2296	0,2327	0,2358	0,2389
-0,7	0,2420	0,2451	0,2483	0,2514	0,2546	0,2578	0,2611	0,2643	0,2676	0,2709
-0,6	0,2743	0,2776	0,2810	0,2843	0,2877	0,2912	0,2946	0,2981	0,3015	0,3050
-0,5	0,3085	0,3121	0,3156	0,3192	0,3228	0,3264	0,3300	0,3336	0,3372	0,3409
-0,4	0,3446	0,3483	0,3520	0,3557	0,3594	0,3632	0,3669	0,3707	0,3745	0,3783
-0,3	0,3821	0,3859	0,3897	0,3936	0,3974	0,4013	0,4052	0,4090	0,4129	0,4168
-0,2	0,4207	0,4247	0,4286	0,4325	0,4364	0,4404	0,4443	0,4483	0,4522	0,4562
-0,1	0,4602	0,4641	0,4681	0,4721	0,4761	0,4801	0,4840	0,4880	0,4920	0,4960
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359

Anexo F. Tabla de la Distribución Normal Estándar (Positiva)

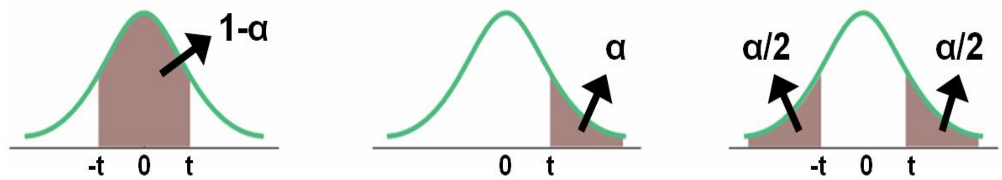
Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Anexo G. Tablas de Distribución Chi-Cuadrado.

	Probabilidad											
Gl	0,95	0,9	0,8	0,7	0,5	0,4	0,3	0,2	0,1	0,05	0,01	0,001
1	0,004	0,02	0,06	0,15	0,45	0,71	1,07	1,64	2,71	3,84	6,63	10,83
2	0,10	0,21	0,45	0,71	1,39	1,83	2,41	3,22	4,61	5,99	9,21	13,82
3	0,35	0,58	1,01	1,42	2,37	2,95	3,66	4,64	6,25	7,81	11,34	16,27
4	0,71	1,06	1,65	2,19	3,36	4,04	4,88	5,99	7,78	9,49	13,28	18,47
5	1,15	1,61	2,34	3,00	4,35	5,13	6,06	7,29	9,24	11,07	15,09	20,52
6	1,64	2,20	3,07	3,83	5,35	6,21	7,23	8,56	10,64	12,59	16,81	22,46
7	2,17	2,83	3,82	4,67	6,35	7,28	8,38	9,80	12,02	14,07	18,48	24,32
8	2,73	3,49	4,59	5,53	7,34	8,35	9,52	11,03	13,36	15,51	20,09	26,12
9	3,33	4,17	5,38	6,39	8,34	9,41	10,66	12,24	14,68	16,92	21,67	27,88
10	3,94	4,87	6,18	7,27	9,34	10,47	11,78	13,44	15,99	18,31	23,21	29,59
11	4,57	5,58	6,99	8,15	10,34	11,53	12,90	14,63	17,28	19,68	24,72	31,26
12	5,23	6,30	7,81	9,03	11,34	12,58	14,01	15,81	18,55	21,03	26,22	32,91
13	5,89	7,04	8,63	9,93	12,34	13,64	15,12	16,98	19,81	22,36	27,69	34,53
14	6,57	7,79	9,47	10,82	13,34	14,69	16,22	18,15	21,06	23,68	29,14	36,12
15	7,26	8,55	10,31	11,72	14,34	15,73	17,32	19,31	22,31	25,00	30,58	37,70
16	7,96	9,31	11,15	12,62	15,34	16,78	18,42	20,47	23,54	26,30	32,00	39,25
17	8,67	10,09	12,00	13,53	16,34	17,82	19,51	21,61	24,77	27,59	33,41	40,79
18	9,39	10,86	12,86	14,44	17,34	18,87	20,60	22,76	25,99	28,87	34,81	42,31
19	10,12	11,65	13,72	15,35	18,34	19,91	21,69	23,90	27,20	30,14	36,19	43,82
20	10,85	12,44	14,58	16,27	19,34	20,95	22,77	25,04	28,41	31,41	37,57	45,31
21	11,59	13,24	15,44	17,18	20,34	21,99	23,86	26,17	29,62	32,67	38,93	46,80
22	12,34	14,04	16,31	18,10	21,34	23,03	24,94	27,30	30,81	33,92	40,29	48,27
23	13,09	14,85	17,19	19,02	22,34	24,07	26,02	28,43	32,01	35,17	41,64	49,73
24	13,85	15,66	18,06	19,94	23,34	25,11	27,10	29,55	33,20	36,42	42,98	51,18
25	14,61	16,47	18,94	20,87	24,34	26,14	28,17	30,68	34,38	37,65	44,31	52,62
26	15,38	17,29	19,82	21,79	25,34	27,18	29,25	31,79	35,56	38,89	45,64	54,05
27	16,15	18,11	20,70	22,72	26,34	28,21	30,32	32,91	36,74	40,11	46,96	55,48
28	16,93	18,94	21,59	23,65	27,34	29,25	31,39	34,03	37,92	41,34	48,28	56,89
29	17,71	19,77	22,48	24,58	28,34	30,28	32,46	35,14	39,09	42,56	49,59	58,30
30	18,49	20,60	23,36	25,51	29,34	31,32	33,53	36,25	40,26	43,77	50,89	59,70
40	26,51	29,05	32,34	34,87	39,34	41,62	44,16	47,27	51,81	55,76	63,69	73,40
50	34,76	37,69	41,45	44,31	49,33	51,89	54,72	58,16	63,17	67,50	76,15	86,66
60	43,19	46,46	50,64	53,81	59,33	62,13	65,23	68,97	74,40	79,08	88,38	99,61
70	51,74	55,33	59,90	63,35	69,33	72,36	75,69	79,71	85,53	90,53	100,43	112,32
80	60,39	64,28	69,21	72,92	79,33	82,57	86,12	90,41	96,58	101,88	112,33	124,84
90	69,13	73,29	78,56	82,51	89,33	92,76	96,52	101,05	107,57	113,15	124,12	137,21
100	77,93	82,36	87,95	92,13	99,33	102,95	106,91	111,67	118,50	124,34	135,81	149,45

Anexo H. Tabla de la Distribución t-Student

Distribucion t de student



Intervalo de confianza | Prueba de una cola | Prueba de dos colas

Grados de libertad												
	Intervalos de confianza de DOS colas (%)											
	50	**60**	**70**	**80**	**85**	**90**	**95**	**96**	**97**	**98**	**99**	**99,9**
	1-α											
	0,5	**0,6**	**0,7**	**0,8**	**0,85**	**0,9**	**0,95**	**0,96**	**0,97**	**0,98**	**0,99**	**0,999**
	Nivel de significancia, α para DOS colas											
	0,5	**0,4**	**0,3**	**0,2**	**0,15**	**0,1**	**0,05**	**0,04**	**0,03**	**0,02**	**0,01**	**0,001**
	Nivel de significancia, α para UNA cola											
	0,25	**0,2**	**0,15**	**0,1**	**0,075**	**0,05**	**0,025**	**0,02**	**0,015**	**0,01**	**0,005**	**0,0005**
1	1,000	1,376	1,963	3,078	4,165	6,314	12,706	15,895	21,205	31,821	63,657	636,619
2	0,816	1,061	1,386	1,886	2,282	2,920	4,303	4,849	5,643	6,965	9,925	31,599
3	0,765	0,978	1,250	1,638	1,924	2,353	3,182	3,482	3,896	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	1,778	2,132	2,776	2,999	3,298	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	1,699	2,015	2,571	2,757	3,003	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,650	1,943	2,447	2,612	2,829	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,617	1,895	2,365	2,517	2,715	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,592	1,860	2,306	2,449	2,634	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,574	1,833	2,262	2,398	2,574	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,559	1,812	2,228	2,359	2,527	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,548	1,796	2,201	2,328	2,491	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,538	1,782	2,179	2,303	2,461	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,530	1,771	2,160	2,282	2,436	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,523	1,761	2,145	2,264	2,415	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,517	1,753	2,131	2,249	2,397	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,512	1,746	2,120	2,235	2,382	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,508	1,740	2,110	2,224	2,368	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,504	1,734	2,101	2,214	2,356	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,500	1,729	2,093	2,205	2,346	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,497	1,725	2,086	2,197	2,336	2,528	2,845	3,850
21	0,686	0,859	1,063	1,323	1,494	1,721	2,080	2,189	2,328	2,518	2,831	3,819
22	0,686	0,858	1,061	1,321	1,492	1,717	2,074	2,183	2,320	2,508	2,819	3,792
23	0,685	0,858	1,060	1,319	1,489	1,714	2,069	2,177	2,313	2,500	2,807	3,768
24	0,685	0,857	1,059	1,318	1,487	1,711	2,064	2,172	2,307	2,492	2,797	3,745
25	0,684	0,856	1,058	1,316	1,485	1,708	2,060	2,167	2,301	2,485	2,787	3,725
26	0,684	0,856	1,058	1,315	1,483	1,706	2,056	2,162	2,296	2,479	2,779	3,707
27	0,684	0,855	1,057	1,314	1,482	1,703	2,052	2,158	2,291	2,473	2,771	3,690
28	0,683	0,855	1,056	1,313	1,480	1,701	2,048	2,154	2,286	2,467	2,763	3,674
29	0,683	0,854	1,055	1,311	1,479	1,699	2,045	2,150	2,282	2,462	2,756	3,659
30	0,683	0,854	1,055	1,310	1,477	1,697	2,042	2,147	2,278	2,457	2,750	3,646
40	0,681	0,851	1,050	1,303	1,468	1,684	2,021	2,123	2,250	2,423	2,704	3,551
50	0,679	0,849	1,047	1,299	1,462	1,676	2,009	2,109	2,234	2,403	2,678	3,496
60	0,679	0,848	1,045	1,296	1,458	1,671	2,000	2,099	2,223	2,390	2,660	3,460
70	0,678	0,847	1,044	1,294	1,456	1,667	1,994	2,093	2,215	2,381	2,648	3,435
80	0,678	0,846	1,043	1,292	1,453	1,664	1,990	2,088	2,209	2,374	2,639	3,416
90	0,677	0,846	1,042	1,291	1,452	1,662	1,987	2,084	2,205	2,368	2,632	3,402
100	0,677	0,845	1,042	1,290	1,451	1,660	1,984	2,081	2,201	2,364	2,626	3,390

Anexo I. Tabla de la Distribución Binomial.

		Probabilidad bruta (no acumulada)										
Ensayos	**Exitos**	**0,01**	**0,05**	**0,1**	**0,15**	**0,2**	**0,25**	**0,3**	**0,35**	**0,4**	**0,45**	**0,5**
1	**0**	0,990	0,950	0,900	0,850	0,800	0,750	0,700	0,650	0,600	0,550	0,500
1	**1**	0,010	0,050	0,100	0,150	0,200	0,250	0,300	0,350	0,400	0,450	0,500
2	**0**	0,980	0,903	0,810	0,723	0,640	0,563	0,490	0,423	0,360	0,303	0,250
2	**1**	0,020	0,095	0,180	0,255	0,320	0,375	0,420	0,455	0,480	0,495	0,500
2	**2**	0,000	0,003	0,010	0,023	0,040	0,063	0,090	0,123	0,160	0,203	0,250
3	**0**	0,970	0,857	0,729	0,614	0,512	0,422	0,343	0,275	0,216	0,166	0,125
3	**1**	0,029	0,135	0,243	0,325	0,384	0,422	0,441	0,444	0,432	0,408	0,375
3	**2**	0,000	0,007	0,027	0,057	0,096	0,141	0,189	0,239	0,288	0,334	0,375
3	**3**	0,000	0,000	0,001	0,003	0,008	0,016	0,027	0,043	0,064	0,091	0,125
4	**0**	0,961	0,815	0,656	0,522	0,410	0,316	0,240	0,179	0,130	0,092	0,063
4	**1**	0,039	0,171	0,292	0,368	0,410	0,422	0,412	0,384	0,346	0,299	0,250
4	**2**	0,001	0,014	0,049	0,098	0,154	0,211	0,265	0,311	0,346	0,368	0,375
4	**3**	0,000	0,000	0,004	0,011	0,026	0,047	0,076	0,111	0,154	0,200	0,250
4	**4**	0,000	0,000	0,000	0,001	0,002	0,004	0,008	0,015	0,026	0,041	0,063
5	**0**	0,951	0,774	0,590	0,444	0,328	0,237	0,168	0,116	0,078	0,050	0,031
5	**1**	0,048	0,204	0,328	0,392	0,410	0,396	0,360	0,312	0,259	0,206	0,156
5	**2**	0,001	0,021	0,073	0,138	0,205	0,264	0,309	0,336	0,346	0,337	0,313
5	**3**	0,000	0,001	0,008	0,024	0,051	0,088	0,132	0,181	0,230	0,276	0,313
5	**4**	0,000	0,000	0,000	0,002	0,006	0,015	0,028	0,049	0,077	0,113	0,156
5	**5**	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,005	0,010	0,018	0,031
6	**0**	0,941	0,735	0,531	0,377	0,262	0,178	0,118	0,075	0,047	0,028	0,016
6	**1**	0,057	0,232	0,354	0,399	0,393	0,356	0,303	0,244	0,187	0,136	0,094
6	**2**	0,001	0,031	0,098	0,176	0,246	0,297	0,324	0,328	0,311	0,278	0,234
6	**3**	0,000	0,002	0,015	0,041	0,082	0,132	0,185	0,235	0,276	0,303	0,313
6	**4**	0,000	0,000	0,001	0,005	0,015	0,033	0,060	0,095	0,138	0,186	0,234
6	**5**	0,000	0,000	0,000	0,000	0,002	0,004	0,010	0,020	0,037	0,061	0,094
6	**6**	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,004	0,008	0,016
7	**0**	0,932	0,698	0,478	0,321	0,210	0,133	0,082	0,049	0,028	0,015	0,008
7	**1**	0,066	0,257	0,372	0,396	0,367	0,311	0,247	0,185	0,131	0,087	0,055
7	**2**	0,002	0,041	0,124	0,210	0,275	0,311	0,318	0,298	0,261	0,214	0,164
7	**3**	0,000	0,004	0,023	0,062	0,115	0,173	0,227	0,268	0,290	0,292	0,273
7	**4**	0,000	0,000	0,003	0,011	0,029	0,058	0,097	0,144	0,194	0,239	0,273
7	**5**	0,000	0,000	0,000	0,001	0,004	0,012	0,025	0,047	0,077	0,117	0,164
7	**6**	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,008	0,017	0,032	0,055
7	**7**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,004	0,008
8	**0**	0,923	0,663	0,430	0,272	0,168	0,100	0,058	0,032	0,017	0,008	0,004
8	**1**	0,075	0,279	0,383	0,385	0,336	0,267	0,198	0,137	0,090	0,055	0,031
8	**2**	0,003	0,051	0,149	0,238	0,294	0,311	0,296	0,259	0,209	0,157	0,109
8	**3**	0,000	0,005	0,033	0,084	0,147	0,208	0,254	0,279	0,279	0,257	0,219
8	**4**	0,000	0,000	0,005	0,018	0,046	0,087	0,136	0,188	0,232	0,263	0,273
8	**5**	0,000	0,000	0,000	0,003	0,009	0,023	0,047	0,081	0,124	0,172	0,219
8	**6**	0,000	0,000	0,000	0,000	0,001	0,004	0,010	0,022	0,041	0,070	0,109
8	**7**	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,003	0,008	0,016	0,031
8	**8**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,004
9	**0**	0,914	0,630	0,387	0,232	0,134	0,075	0,040	0,021	0,010	0,005	0,002
9	**1**	0,083	0,299	0,387	0,368	0,302	0,225	0,156	0,100	0,060	0,034	0,018
9	**2**	0,003	0,063	0,172	0,260	0,302	0,300	0,267	0,216	0,161	0,111	0,070
9	**3**	0,000	0,008	0,045	0,107	0,176	0,234	0,267	0,272	0,251	0,212	0,164
9	**4**	0,000	0,001	0,007	0,028	0,066	0,117	0,172	0,219	0,251	0,260	0,246
9	**5**	0,000	0,000	0,001	0,005	0,017	0,039	0,074	0,118	0,167	0,213	0,246
9	**6**	0,000	0,000	0,000	0,001	0,003	0,009	0,021	0,042	0,074	0,116	0,164
9	**7**	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,010	0,021	0,041	0,070
9	**8**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,008	0,018
9	**9**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002
10	**0**	0,904	0,599	0,349	0,197	0,107	0,056	0,028	0,013	0,006	0,003	0,001
10	**1**	0,091	0,315	0,387	0,347	0,268	0,188	0,121	0,072	0,040	0,021	0,010
10	**2**	0,004	0,075	0,194	0,276	0,302	0,282	0,233	0,176	0,121	0,076	0,044
10	**3**	0,000	0,010	0,057	0,130	0,201	0,250	0,267	0,252	0,215	0,166	0,117
10	**4**	0,000	0,001	0,011	0,040	0,088	0,146	0,200	0,238	0,251	0,238	0,205
10	**5**	0,000	0,000	0,001	0,008	0,026	0,058	0,103	0,154	0,201	0,234	0,246
10	**6**	0,000	0,000	0,000	0,001	0,006	0,016	0,037	0,069	0,111	0,160	0,205
10	**7**	0,000	0,000	0,000	0,000	0,001	0,003	0,009	0,021	0,042	0,075	0,117
10	**8**	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,004	0,011	0,023	0,044
10	**9**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,002	0,004	0,010
10	**10**	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001

Joan Fernando Chipia Lobo.

- Profesor Ordinario de Bioestadística, categoría Asistente, perteneciente al Departamento de Medicina Preventiva y Social, Facultad de Medicina, Universidad de Los Andes (Mérida, Venezuela).
- Licenciado en Educación mención Matemática, Universidad de Los Andes (ULA).
- Magíster en Educación mención Informática y Diseño Instruccional, ULA.
- Magíster en Salud Pública, ULA.
- Diplomado en Informática Educativa, ULA.
- Diplomado internacional en Derechos del Niño Asociación de Naciones Unidas de Venezuela (ANUV) y Universidad Pedagógica Experimental Libertador (UPEL).
- Diplomado internacional en Tecnologías de la Información y Comunicación, ANUV y UPEL.
- Curso Medio de Salud Pública, ULA.
- Curso Básico Docente Básico en Educación Superior, ULA.
- Diplomado en Estadística Computarizada y Análisis de Datos, ULA.
- Ponente y facilitador en eventos académicos nacionales e internacionales.
- Investigador acreditado por ULA desde 2011 y por ONCTI desde 2012.
- Editor en Jefe de la Revista GICOS.
- Doctorando en Gerencia Avanzada, Universidad Fermín Toro.