Universidad
Continental

3i INNOVADOR INTERACTIVO INTEGRAL

Probabilidad y Estadística

José Antonio Baldeón Crisóstomo

MAI

Datos de catalogación bibliográfica

Probabilidad y Estadística. Manual Autoformativo Interactivo
José Antonio Baldeón Crisóstomo

Primera edición
Huancayo, agosto de 2016

De esta edición
© Universidad Continental
Av. San Carlos 1980, Huancayo-Perú
Teléfono: (51 64) 481-430 anexo 7361
Correo electrónico: recursosucvirtual@continental.edu.pe
http://www.continental.edu.pe/

Versión e-book
Disponible en http://repositorio.continental.edu.pe/
ISBN electrónico N.° 978-612-4196-

Dirección: Emma Barrios Ipenza
Edición: Eliana Gallardo Echenique
Asistente de edición: Andrid Poma Acevedo
Asesoría didáctica: Karine Bernal
Corrección de textos: Eliana Gallardo Echenique
Diseño y diagramación: Francisco Rosales Guerra

ÍNDICE

UNIDAD IV “DISTRIBUCIONES DE PROBABILIDAD” 117

INTRODUCCIÓN

Cuando escuchamos el término ESTADÍSTICA lo relacionamos frecuentemente con el término datos o datos individuales y porcentajes o cualquier información relacionada con ellos. Sin embargo la ESTADÍSTICA es más que eso, no es sólo una serie de datos o un gráfico estadístico, involucra varias etapas que son materia de estudio del presente manual autoformativo.

Podemos decir que la función principal de la estadística es justamente la recolección y agrupamiento de datos de diverso tipo para construir con ellos informes estadísticos que nos den idea sobre diferentes y muy variados temas, siempre desde un punto de vista cuantitativo y no cualitativo. Esto es muy importante remarcarlo ya que la estadística se convierte entonces en una ciencia que nos habla de cantidades (por ejemplo, cuántas personas viven en un país por metro cuadrado) pero no nos da información directa sobre la calidad de vida de esas personas.

Este análisis numérico que se realiza con los datos es muy importante puesto que con ello fundamentaremos nuestra toma de decisiones, desde una simple actividad diaria de alguna persona hasta una complicada gestión en alguna empresa, en cualquier campo de acción. Es por ello que se requiere del conocimiento para poder disponer de datos apropiados, suficientes, oportunos y de buena calidad, así como manejar una buena metodología y manejo de procesos que nos permitan alcanzar el conocimiento deseado acerca de la realidad para realizar una buena toma de decisiones. No hay que dejar de ver que la ESTADÍSTICA al brindar apoyo a otras ciencias requiere del buen manejo de procesos dirigidos a la toma y análisis de datos asociándolos con el buen manejo del cálculo de las probabilidades y con ciertos niveles de confiabilidad en la toma de decisiones.

El presente Manual Autoformativo de PROBABILIDAD Y ESTADÍSTICA está diseñado para que el lector pueda adquirir los conocimientos necesarios para poder ejecutar los procesos relacionados con la descripción y análisis de datos (ESTADÍSTICA DESCRIPTIVA) y el cálculo de probabilidades, utilizando con total autonomía los conceptos, características y ejemplos sobre el Resumen y Gráfica de datos descritos en la Primera Unidad; Análisis Exploratorio de datos e Indicadores Estadísticos descritos en la Segunda Unidad; Probabilidad descrita en la tercera Unidad y Distribuciones de Probabilidad descritas en la Cuarta Unidad; complementadas en cada unidad con lecturas seleccionadas, actividades y autoevaluaciones del aprendizaje.

Agradecemos a quienes de antemano colaboraron con paciencia y comprensión en la elaboración del presente manual y a ustedes estimados(as) estudiantes les deseamos éxitos en el desarrollo de la asignatura.

El Autor

DIAGRAMA DE PRESENTACIÓN DE LA ASIGNATURA

RESULTADOS DEL APRENDIZAJE:

Al finalizar la asignatura, el estudiante será capaz de interpretar la información haciendo uso de los métodos y técnicas de la estadística descriptiva y la teoría de probabilidades relacionados a su profesión.

UNIDADES DIDACTICAS:

UNIDAD I	UNIDAD II	UNIDAD III	UNIDAD IV
Introducción a la estadística. Tipos de distribuciones y gráfica de datos.	Estadísticos para describir, explorar y comparar datos.	Probabilidad	Distribuciones de probabilidad: Distribuciones de probabilidad discreta. Distribuciones de probabilidad normal.

TIEMPO MINIMO DE ESTUDIO:

UNIDAD I	UNIDAD II	UNIDAD III	UNIDAD IV
1ra. y 2da. Semana (16 horas)	1ra. y 2da. Semana (16 horas)	1ra. y 2da. Semana (16 horas)	1ra. y 2da. Semana (16 horas)

UNIDAD I

INTRODUCCIÓN A LA ESTADÍSTICA.TIPOS DE DISTRIBUCIONES Y GRÁFICA DE DATOS.

DIAGRAMA DE PRESENTACIÓN DE LA UNIDAD I

RESULTADO DE APRENDIZAJE

Al finalizar la unidad, el estudiante será capaz de interpretar los conceptos bá-sicos de la estadística y las distribuciones unidimensionales y bidimensionales respetando las leyes y propiedades de la estadística

CONTENIDOS	ACTIVIDADES FORMATIVAS (HABILIDADES Y ACTITUDES)	SISTEMA DE EVALUACIÓN (TÉCNICAS Y CRITERIOS)
TEMA Nº 1: INTRODUCCIÓN A LA ESTADÍSTICA: 1 Estadística, división, objetivos, definiciones, tipos de datos. 2 Fuentes y Métodos de recolección de datos. **TEMA Nº 2: TIPOS DE DISTRIBUCIONES Y GRÁFICA DE DATOS:** 1 Distribuciones de Frecuencias 1.1. Definiciones Básicas 1.2. Distribución de frecuencias unidimensionales (1 variable) 1.3. Distribución de frecuencias bidimensionales (2 variable) 1.4. Gráficos estadísticos 1.5. Principales tipos de gráficos 1.6. Histograma , polígono de frecuencias y ojiva "Menor que " 1.7. Diagrama de Pareto	· Indaga y analiza los conceptos básicos de Estadística. · Identifica y utiliza los métodos y las fuentes de recolección de datos. · Mide, describe, explora y compara diferentes características de un conjunto de datos. · Construye gráficos estadísticos y analiza e interpreta los resultados. · Resuelve todas las actividades del material de estudio, haciendo uso del laboratorio de cómputo.	Procedimientos e indicadores de evaluación permanente · Entrega puntual de trabajos realizados. · Calidad, coherencia y pertinencia de contenidos desarrollados. · Prueba teórico-práctica, individual. · Actividades desarrolladas en sesiones tutorizadas Criterios de evaluación para la encuesta elaborada: · El tema de la encuesta corresponde a la carrera profesional del estudiante. · Las preguntas corresponden a diferentes aspectos del tema tratado y están basadas en un marco teórico. · La redacción es impecable, sin errores ortográficos, con la correcta expresión. · La encuesta cuenta con: Instrucciones, diferentes tipos de preguntas, variables en diferentes niveles de medición, etiquetas en donde sea necesario, agradecimiento.

RECURSOS:

Videos:

Tema Nº 1

Etapas de una investigación estadística

https://www.youtube.com/watch?v=fRvL6WGEF9U

Tema Nº 2

¿Cómo hacer una encuesta?

https://www.youtube.com/watch?v=q3Rgwxx3il4

Lectura complementaria:

Lectura Seleccionada Nº 1

Seis grados de Kevin Bacon: ¿El estudio original utilizó buenos datos?

Triola, M.(2009). *Estadística.* México: Pearson Educación.

Lectura Seleccionada Nº 2

¿Los Premios de la Academia discriminan por la edad? ¿El estudio original utilizó buenos datos?

Mario Triola – (2009). México: Pearson Educación.

INSTRUMENTO DE EVALUACIÓN	Prueba de desarrollo
BIBLIOGRAFÍA (BÁSICA Y COMPLEMENTARIA)	**BÁSICA** Devore, J. (2008). *Probabilidad y Estadística* (7 Ed). México: Editorial Cengage Learning Editores S.A. Martínez Bencardino, C. (2012). *Estadística y muestreo.* (13 ed). Bogotá, Colombia: Editorial Ecoe Ediciones. Toma, J. (2012). *Estadística aplicada – Primera parte.* (2 Ed). Perú: Fondo editorial de la Universidad del Pacífico **COMPLEMENTARIA** Ross, S. (2001). *Probabilidad y Estadística para Ingenieros.* (3 ed). México: Mc Graw Hill. Berenson, M. & Levine, D. (2010). *Estadística Básica en Administración, Conceptos y aplicaciones.* México : Prentice Hall.
RECURSOS EDUCATIVOS DIGITALES	· Instituto Nacional de Estadística e Informática. Disponible en: http://www.inei.gob.pe/ · Tokio explora en un congreso la "imprescindible" ciencia de la estadística. (2012, Jul 04). EFE News Service. Disponible en: http://search.proquest.com/docview/1023206430?accountid=146219 · Canales, E. (2005, Jul 26). Mexicar / AMLO sin estadistica. El Norte. Disponible en:: http://search.proquest.com/docview/311786309?accountid=146219

TEMA Nº 1:
INTRODUCCIÓN A LA ESTADÍSTICA.

Estimado estudiante, en esta primera unidad conoceremos los aspectos y conceptos básicos de la disciplina Estadística, los cuales serán de vital importancia al momento de llevar a cabo los diferentes procedimientos que estudiaremos más adelante en nuestra asignatura. El propósito de esta unidad es que logres diferenciar entre lo que es población, parámetro, muestra, estadístico, variable, tipo de variable, datos, nivel de medición. El éxito que consigas dependerá mucho de que hayas internalizado bien estos primeros temas.

1. ESTADÍSTICA

Es la ciencia que le facilita al hombre el estudio de datos masivos, proporcionando un conjunto de métodos científicos para recolectar, resumir, clasificar, analizar e interpretar el comportamiento de los datos con respecto a una característica, materia de estudio o investigación, pasa de esa manera a sacar conclusiones valederas y efectuar predicciones razonables de ellos y así mostrar una visión de conjunto clara y de más fácil apreciación con respecto a la fuente de información, que nos permiten tomar decisiones óptimas en casos de incertidumbre

1.1. División de la Estadística:

Estadística descriptiva o deductiva: Se emplea para describir las características básicas de los datos en estudio. Por medio de gráficas y valores numéricos calculados algebraicamente, proporciona resúmenes simples de los datos de que dispone. Simplemente se describe lo que son o lo que muestran los datos y no pretende ni implica ir más allá de los datos de que se dispone. Por tanto, puede decirse que la Estadística descriptiva es una serie de métodos para organizar, resumir y presentar datos de manera informativa.

Estadística inferencial o inductiva: Se emplea para extraer algún tipo de conclusión (hacer una inferencia de alguna clase) y tomar decisiones respecto a la población con base en la información obtenida mediante la observación de relativamente pocos elementos individuales de la población (muestra). Es decir, la muestra es un medio para llegar a un fin en lugar de un fin por sí misma. Por tanto, puede decirse que la Estadística inferencial está formada por las técnicas para generalizar desde una muestra hasta una población.

1.2. Objetivos de la Estadística

Los objetivos de la estadística pueden ser clasificados en tres grandes capítulos: Descripción, análisis y predicción.

- Descripción de grandes colecciones de datos empíricos reduciéndolos a un pequeño número de características que concentra la parte más importante y significativa de la información proporcionada por los datos.

  La descripción supone que los datos que vienen expresados en su forma natural deben ser clasificados y presentados sistemáticamente en cuadros o tablas como una pequeña reducción de datos, esto se obtiene cuando el comportamiento y características de los datos se expresan por un conjunto de indicadores, medidas de resumen o estadígrafos.

- Análisis estadístico de datos experimentales y de los fenómenos observados. Toda la investigación estadística incluye un problema de análisis, con el objeto de formarse un concepto de la población o universo y adoptar decisiones; en este caso no es necesario observar a toda una población sino que será suficiente elegir una muestra representativa. La preocupación del análisis estadístico es inferir propiedades para una población sobre la base de resultados muestrales conocidos.

Todo análisis debe suponer la elección adecuada de una muestra representativa, la que será estudiada en detalle para obtener conclusiones o resultados, que dentro de ciertos márgenes de aceptación sean válidos para toda la población de la cual fue elegida la muestra.

- Predicción o comportamiento de los fenómenos en el futuro, lo cual constituye la máxima aspiración práctica de toda ciencia. Este objetivo de predicción y previsión está implícito tanto en la descripción como en el análisis estadístico, puesto que en general interesa orientar la toma de decisiones con vigencia y efecto en el futuro. Naturalmente que las estimaciones y proyecciones dependen del grado de conocimiento del comportamiento del pasado y presente de las variables en estudio.

1.3. Definiciones básicas:

Población: Es una colección bien definida de todos los objetos o elementos (puntajes. personas, mediciones, etc.) que interesan a un cierto estudio.

La población debe estar perfectamente definida en el tiempo y en el espacio, de modo que ante la presencia de un potencial integrante de la misma, se pueda decidir si forma parte o no de la población bajo estudio. Por lo tanto, al definir una población, se debe cuidar que el conjunto de elementos que la integran quede perfectamente delimitado. La población puede ser según su tamaño de dos tipos:

Población finita: Cuando se tiene un número determinado de elementos. Ejemplo: La población conformada por los estudiantes de la Universidad "Continental" de la ciudad de Huancayo.

Población infinita: Cuando el número de elementos es indeterminado, o tan grande que pudiesen considerarse infinitos. Ejemplo: La población formada por todas las hormigas que habitan las riberas del río Amazonas.

Nota:

- El concepto de población no necesariamente se refiere a personas.
- Cuando la información deseada está disponible para todos los objetos de la población, se tiene lo que se llama un "Censo". Las restricciones de tiempo, dinero y otros recursos escasos casi siempre hacen que un censo no sea factible.

Muestra: Subconjunto o parte de una población que se selecciona con la finalidad de analizar las características y propiedades de la población de la cual ha sido tomada. Es deseable que la muestra sea representativa, lo cual significa que debe tener las mismas características que se manifiestan en la población de la que proviene. Una buena muestra sería como una versión a escala de la población, que reflejaría cada una de las características de ésta. Por supuesto una muestra perfecta como ésta no existe para poblaciones complejas.

Para seleccionar una muestra se puede recurrir a la teoría del muestreo, que sostiene que si los elementos de análisis u observación son seleccionados en forma aleatoria, es decir, al azar (sin preferencia alguna), se puede trabajar con base en muestras para tener un conocimiento con un cierto nivel de confianza de ciertas características de la población. Cuando no se emplea un mecanismo aleatorio para seleccionar la muestra, sino que sea hace con base en criterios de conveniencia, se obtienen resultados que pueden ser útiles. Sin embargo, no hay una teoría que pueda desarrollarse y respaldar tales resultados.

Nota**:**

- El proceso de recolección de datos muestrales se llama "Muestreo".

Parámetro: Es una cantidad numérica calculada sobre una población y resume los valores que ésta toma en algún atributo. Intenta resumir toda la información que hay en la población en unos pocos números. Se considera como un valor verdadero de la característica estudiada y para determinar su valor es necesario utilizar la información poblacional completa, y por lo tanto, la decisión se toma con certidumbre total. Se le representa con letras griegas en minúscula (m, s, p). Generalmente los parámetros no son conocidos o quizá nunca se lleguen a conocer.

Estadístico o estadígrafo:

Es una cantidad numérica calculada sobre una muestra que resume su información sobre algún aspecto. Intenta resumir la información de la muestra en unos pocos números. Normalmente nos interesa conocer un parámetro, pero por la dificultad que conlleva estudiar a toda la población, calculamos un estadígrafo sobre una muestra y "confiamos" en que sean próximos.

Se le representa con letras latinas en minúscula (x, s, p).

Nota**:**

- Si un estadístico se usa para aproximar un parámetro también se le suele llamar "estimador".

Variable: Es cualquier característica cuyo valor puede cambiar de un objeto a otro en la población. Se denotan con las letras minúsculas del alfabeto.

Las variables se pueden clasificar de la siguiente manera:

- Cualitativa: También llamada de atributo o categórica. Es aquella variable de naturaleza no numérica. Por lo tanto, con sus valores no se pueden realizar operaciones aritméticas.

Ejemplo: Género, estado civil, lugar de nacimiento.

- Cuantitativa: Es aquella variable de naturaleza numérica. Con sus valores se puede efectuar operaciones aritméticas. A su vez, se clasifica en:
  - Variable cuantitativa discreta: Es aquella variable cuantitativa que puede tomar sólo valores enteros. Generalmente se obtienen por conteo.

    Ejemplo: Número de hijos en una familia, cantidad de goles en un partido de fútbol, número de viviendas de un centro poblado.
  - Variable cuantitativa continua: Es aquella variable cuantitativa que puede tomar cualquier valor en un cierto rango de valores. Generalmente se obtienen por medición.

    Ejemplo: Peso, estatura, temperatura.

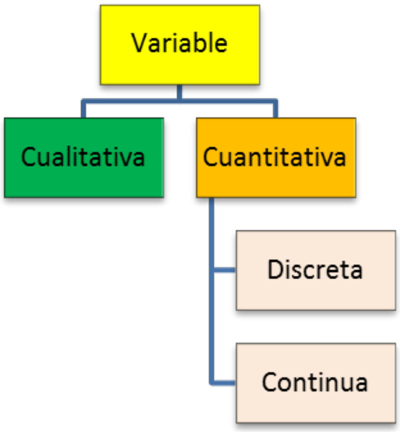


Figura 1. Tipos de variable
Fuente: Elaboración propia

Dato: Es el valor, respuesta o registro que adquiere una característica o variable asociado a un elemento de la población o muestra, como resultado de la observación, entrevista o recopilación en general. Puede ser un número, una palabra o un símbolo. Los datos son el insumo de la Estadística.

Elementos o unidades estadísticas: Son las entidades acerca de las cuales se recogen los datos. Pueden ser personas, objetos o cosas. Asimismo, pueden ser entidades simples (una persona) o entidades complejas (una familia).

Niveles o escalas de medición: Los datos se clasifican por niveles de medición. El nivel de medición de los datos rige los cálculos que se llevan a cabo con el fin de resumir y presentar los datos. También determina las pruebas estadísticas que se deben realizar. Pueden ser:

- **Nivel nominal:** Las observaciones acerca de una variable se caracterizan por datos que consisten exclusivamente en nombres, etiquetas o categorías. Los datos no se pueden acomodar en un esquema de orden (como del más bajo al más alto).

  Ejemplo: Los datos de la variable Género: Masculino y femenino.

- **Nivel ordinal:** Cuando los datos pueden acomodarse en algún orden, aunque las diferencias entre los valores de los datos (obtenidas por medio de una resta) no pueden calcularse o carecen de significado. Implica jerarquía entre los datos.

  Ejemplo: Los datos de la variable Grado de instrucción: Analfabeto, primaria, secundaria, superior.

- **Nivel de intervalo:** Se parece al nivel ordinal, pero con la propiedad adicional de que la diferencia entre dos valores de datos cualesquiera tiene un significado. Sin embargo, los datos en este nivel no tienen punto de partida cero natural inherente (donde la cantidad que está presente corresponde a nada).

  Ejemplo: Los datos de la variable Temperatura.

- **Nivel de razón:** Es similar al nivel de intervalo, pero con la propiedad adicional de que sí tiene un punto de partida cero natural (donde el cero indica que nada de la cantidad está presente). Para valores en este nivel, tanto las diferencias como las razones tienen significado.

  Ejemplo: Los datos de la variable Peso.

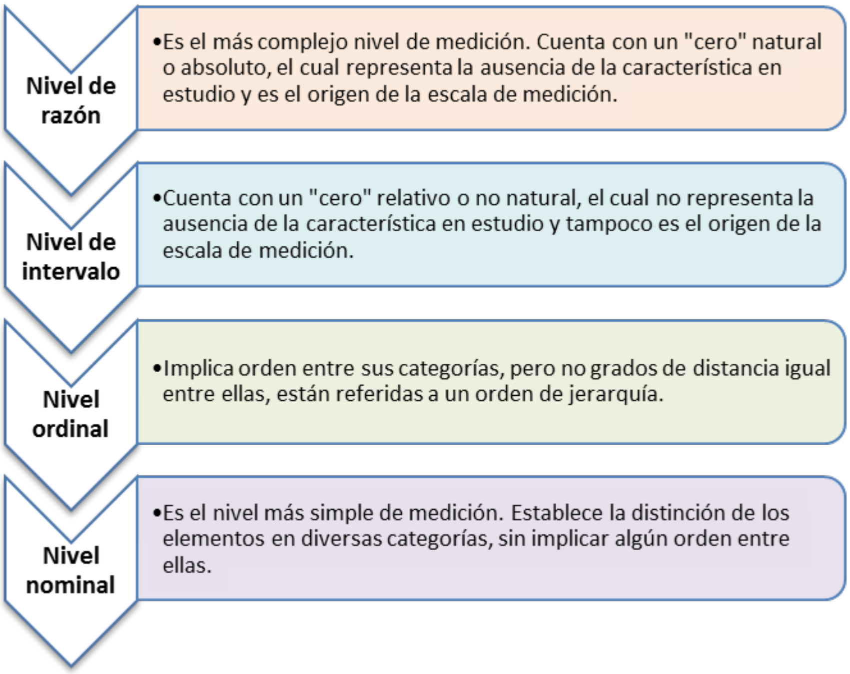


Figura 2. Niveles de medición
Fuente: Elaboración propia

2. FUENTES Y MÉTODOS DE RECOLECCIÓN DE DATOS

2.1. FUENTES DE DATOS

Las fuentes de información están constituidas por cada uno de los lugares de donde se toman los datos. De acuerdo al tipo del lugar del cual procede esta información podemos clasificarlo de la siguiente manera:

Fuente Primaria

Los datos de la fuente primaria son obtenidos directamente de las unidades de observación mediante cualquier técnica o instrumento de recolección de datos originales.

Fuente Secundaria

Los datos de la fuente secundaria son aquellos datos que ya han sido publicados con anterioridad, recolectados con fines diferentes de los que la investigación específica necesita.

Estos datos se encuentran como archivos, registros administrativos, boletines, informes estadísticos requeridos en el ámbito nacional o sectorial; elaborados por organismos especializados, los que pueden ser públicos o privados.

Fuentes Internas

Los datos procedentes de fuentes internas son aquellos que se generan dentro de la propia organización, por ejemplo, productos o servicios ofrecidos, el número de horas de trabajo consumidas en cada unidad de producción, la cantidad de materiales utilizados o desperdiciados, el número de ausencias al trabajo, etc.

Fuentes Externas

Son los datos que se generan fuera de los negocios u organizaciones. Entre las fuentes más importantes de esta clase de información están las agencias gubernamentales, las asociaciones profesionales y comerciales, las publicaciones especializadas de empresas privadas.

2.2. MÉTODOS DE RECOLECCIÓN DE DATOS

Elegir el método de recolección de datos depende de las posibilidades de acceso o contacto con los elementos investigados, del tamaño de la población o muestra, de la oportunidad de obtener datos y del presupuesto y exigencias del tiempo.

Los objetivos principales para la recolección son:

- Obtener los datos o respuestas a las variables analizadas.
- Proporcionar información adecuada y oportuna con fines de una óptima planificación.

Para seleccionar el método de recolección de datos se debe tener en cuenta lo siguiente:

- Establecer objetivos claros: Antes de recoger la información se debe decidir qué se va a hacer con ella. Cualquier recolección de información ha de tener un objetivo específico y ser seguida por acciones.
- Definir su propósito: Una vez que se define el objetivo de la recolección de la información, también se determina los tipos de comparación que se necesitan, y esto a su vez identifica el tipo de datos que se deben de recoger.

- Confiabilidad de las mediciones: Está directamente relacionada a la adecuada selección de la muestra.

CENSO

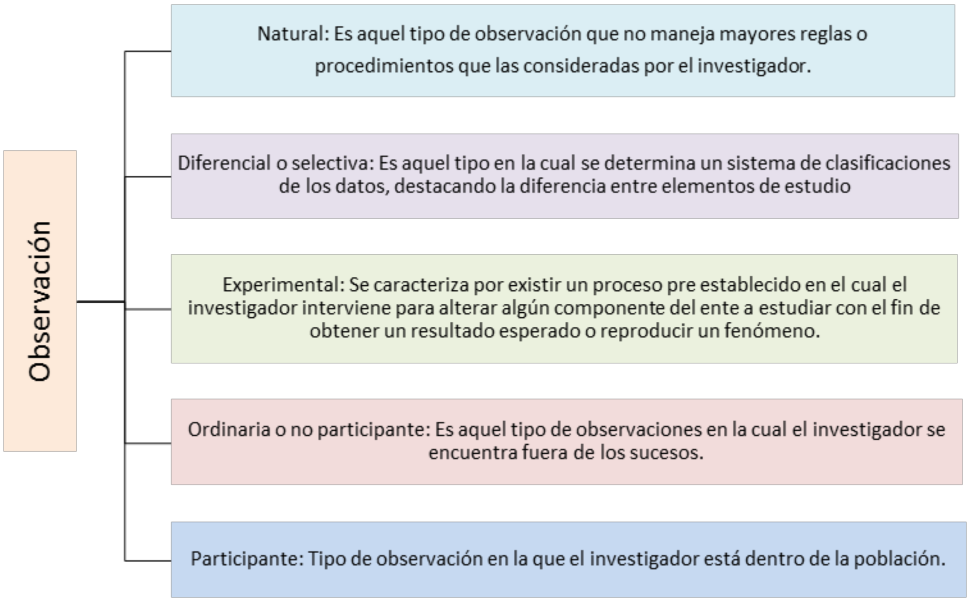
Es un método de recolección de datos mediante el cual la información se obtiene analizando a la totalidad de los elementos que componen la población o universo bajo estudio. Un censo debe cumplir las condiciones de universalidad (censar a todos los elementos de la población) y simultaneidad (realizarse en un momento determinado). Un censo es equivalente a una fotografía de la población bajo estudio.

OBSERVACIÓN

Es un proceso permanente de investigación realizado con instrumentos y técnicas específicas según el ámbito de estudio. Es necesario que el investigador cuente con un marco teórico y referencial sobre las variables y sus indicadores. Para lograr una observación científicamente válida se debe:

- Preparar la estrategia de obtención de información verídica.
- Determinar controles de las variables.
- Planificación y determinación de método y tipo de observación.

Se pueden considerar los siguientes tipos de observación:



Natural: Es aquel tipo de observación que no maneja mayores reglas o procedimientos que las consideradas por el investigador.

Diferencial o selectiva: Es aquel tipo en la cual se determina un sistema de clasificaciones de los datos, destacando la diferencia entre elementos de estudio

Experimental: Se caracteriza por existir un proceso pre establecido en el cual el investigador interviene para alterar algún componente del ente a estudiar con el fin de obtener un resultado esperado o reproducir un fenómeno.

Ordinaria o no participante: Es aquel tipo de observaciones en la cual el investigador se encuentra fuera de los sucesos.

Participante: Tipo de observación en la que el investigador está dentro de la población.

Figura 3. Tipos de Observación
Fuente: Elaboración propia

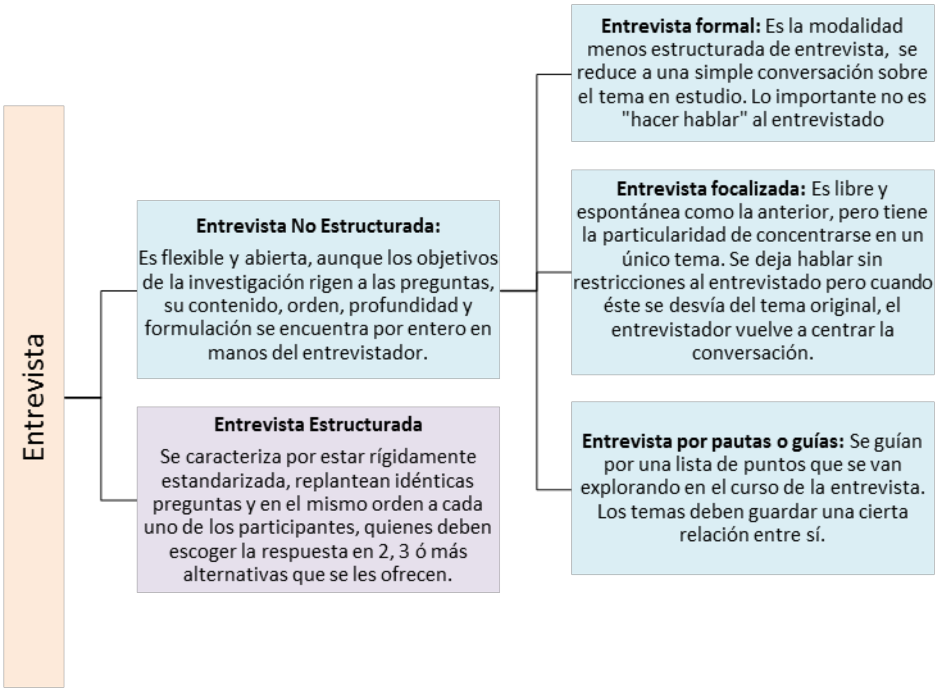
ENTREVISTA

Consiste en una interacción entre dos personas, el entrevistador (quien investiga) que formula una serie de preguntas relativas al tema de investigación, y el entrevistado (quien tiene la información) que responde verbalmente o por escrito las preguntas que le son formuladas.

Esta técnica se aplica a informantes claves, es decir, personas representativas que manejan una gran cantidad de información referente a un tema de interés.

La ventaja de realizar una entrevista es que la información que se obtiene puede ser enriquecida con repreguntas, la observación directa permite constatar la veracidad de las respuestas. Al ser una interacción entre dos personas existen muchos factores que pueden distorsionar las respuestas del entrevistado por lo que el entrevistador debe ser una persona con mucho dominio de esta técnica.

Según la estructura de la entrevista, estas se clasifican en:



Entrevista No Estructurada: Es flexible y abierta, aunque los objetivos de la investigación rigen a las preguntas, su contenido, orden, profundidad y formulación se encuentra por entero en manos del entrevistador.

Entrevista Estructurada Se caracteriza por estar rígidamente estandarizada, replantean idénticas preguntas y en el mismo orden a cada uno de los participantes, quienes deben escoger la respuesta en 2, 3 ó más alternativas que se les ofrecen.

Entrevista formal: Es la modalidad menos estructurada de entrevista, se reduce a una simple conversación sobre el tema en estudio. Lo importante no es "hacer hablar" al entrevistado

Entrevista focalizada: Es libre y espontánea como la anterior, pero tiene la particularidad de concentrarse en un único tema. Se deja hablar sin restricciones al entrevistado pero cuando éste se desvía del tema original, el entrevistador vuelve a centrar la conversación.

Entrevista por pautas o guías: Se guían por una lista de puntos que se van explorando en el curso de la entrevista. Los temas deben guardar una cierta relación entre sí.

Figura 4

Fuente: Elaboración propia

ENCUESTA

Permite obtener información de una muestra representativa de una determinada población. Es un proceso a través del cual conseguimos datos de primera mano y todos ellos que permitan especificar mejor el problema.

CUESTIONARIO

Es un plan formalizado para recolectar datos de los encuestados. La función del cuestionario es la medición del comportamiento pasado, de las actitudes y de las características del encuestado.

Es el método que utiliza un instrumento o formulario impreso, destinado a obtener repuestas sobre el problema en estudio y que el investido o consultado llena por sí mismo.

El cuestionario puede aplicarse a grupos o individuos estando presente el investigador o el responsable del recoger la información, o puede enviarse por correo a los destinatarios seleccionados en la muestra.

Algunas ventajas del cuestionario son: Su costo relativamente bajo, su capacidad para proporcionar información sobre un mayor número de personas en un periodo bastante breve y la facilidad de obtener, cuantificar, analizar e interpretar los datos.

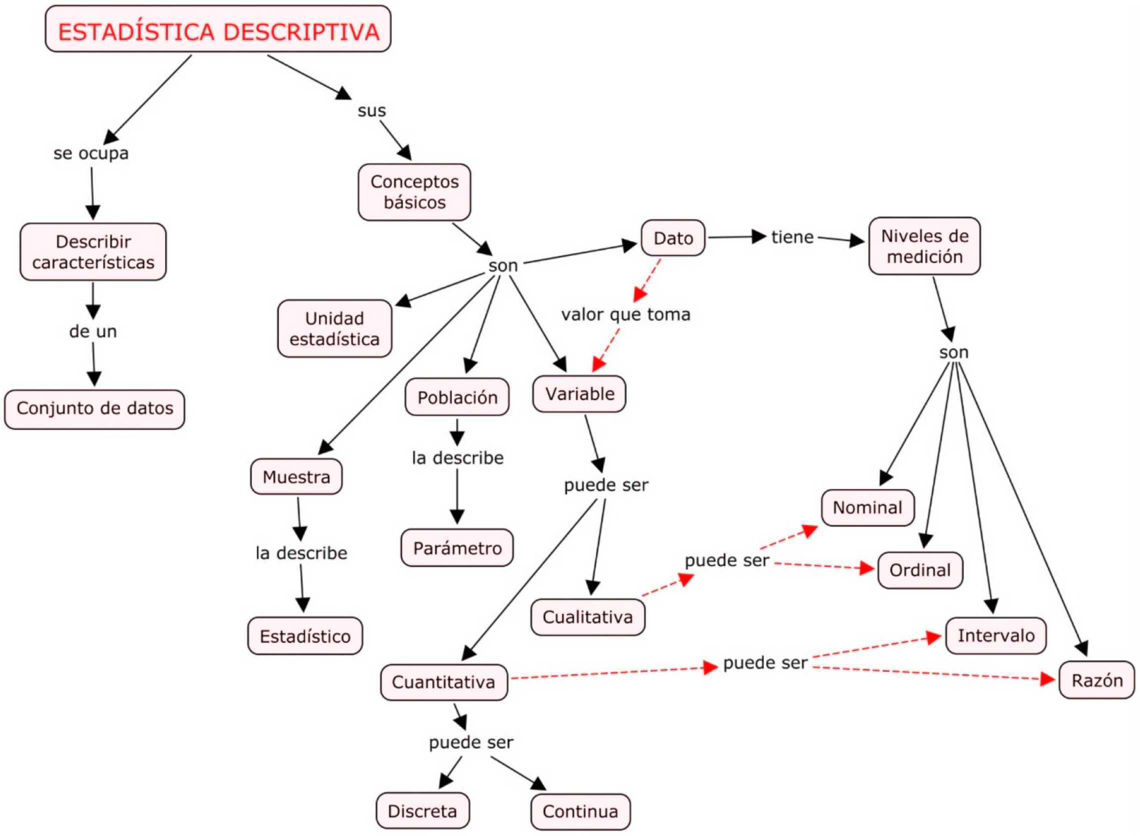
Dentro de las limitaciones de este método figuran las siguientes: Es poco flexible, la información no puede variar ni profundizarse.

Respecto al diseño del cuestionario, éste varía según la experiencia del investigador, los objetivos a alcanzar, los tiempos de aplicación, el presupuesto con que se cuenta, el tiempo para el estudio, entre otros.

Generalmente un cuestionario tiene cinco secciones:

- **Solicitud de cooperación:** Es un pequeño enunciado, diseñado para obtener la cooperación del encuestado con relación a la entrevista, Contiene la identificación de la organización que realiza la encuesta, se explica el objeto del estudio y se indica el tiempo que se requiere para completar la entrevista.
- **Datos de identificación:** generalmente ocupan la primera sección del cuestionario y se relacionan con el nombre, dirección y número telefónico del encuestado. Los datos adicionales incluirían elementos tales como la hora y la fecha de la entrevista, además del nombre o código del entrevistador.
- **Datos de clasificación:** Tratan sobre las características del encuestado. Estos datos los suministra directamente el encuestado en el caso de una encuesta por correo. En las personales y telefónicas el entrevistador recolecta los datos o, en algunos casos, puede estimar tipos más sensibles de datos basado en la observación.
- **Instrucciones:** Se refieren a comentarios realizados al entrevistador o encuestado con relación a la forma de utilizar el cuestionario. Estos comentarios aparecen directamente en el cuestionario cuando se emplea una encuesta por correo.
- **Información solicitada**: Constituye la parte más grande del cuestionario.

EN SÍNTESIS:



Fuente: Elaboración propia

LECTURA SELECCIONADA N.º 1:

Leer apartado: Seis grados de Kevin Bacon: ¿El estudio original utilizó buenos datos?

Triola, M.(2009). *Estadística.* México: Pearson Educación.

"Seis grados de Kevin Bacon" es un juego popular reciente, que consiste en identificar a un actor o a una actriz de cine, y luego vincularlo con el actor Kevin Bacon. (En el momento en que se escribió esto, el juego podía jugarse en el sitio Web www.cs.virginia.edu/oracle). Consideremos a Richard Gere como ejemplo. Gere actuó en la película *Cotton Club* con Laurence Fishburne, que trabajó en la película *Mystic River* con Kevin Bacon. El vínculo Gere-Fishburne-Bacon tiene *dos* grados de separación porque no se cuenta la persona meta.

Este juego, creado por tres estudiantes (Craig Fass, Brian Turtle y Mike Ginelli) de Albright College, es una versión más especializada de "Small World Problem", que plantea la siguiente pregunta: ¿Cuántos intermediarios (amigos, parientes y otros conocidos) se necesitan para conectar a cualesquiera dos personas elegidas al azar en la Tierra? Es decir, para cualesquiera dos personas en nuestro planeta, ¿cuál es el número de grados de separación? Este problema de conexión tiene aplicaciones prácticas en muchos campos, como las redes eléctricas, el uso de Internet, las neuronas del cerebro y la propagación de enfermedades.

El concepto de "seis grados de separación" surgió de un estudio realizado en 1967 por el psicólogo Stanley Milgram, quien originalmente describió que en Estados Unidos dos residentes al azar están conectados por un promedio de seis intermediarios. En su primer experimento, Milgram envió 60 cartas a personas de Wichita, Kansas, a quienes les pidió que renviaran esas cartas a una mujer específica en Cambridge, Massachusetts. A esas personas se les dio la instrucción de entregar en mano las cartas a conocidos que, según ellos, podrían contactar a la persona indicada, ya fuera directamente o a través de otros conocidos. Participaron 50 de las 60 personas, y tres cartas llegaron a su destino. Dos experimentos posteriores tuvieron tasas de terminación más bajas; pero finalmente Milgram alcanzó una tasa del 35 por ciento, y describió que cada cadena completa tenía un promedio de alrededor de seis intermediarios. Como consecuencia, los datos originales de Milgram produjeron el concepto "seis grados de separación".

Veamos dos preguntas clave: ¿Eran adecuados los datos originales de Milgram? ¿Los datos originales de Milgram justifican el concepto de "seis grados de separación?"

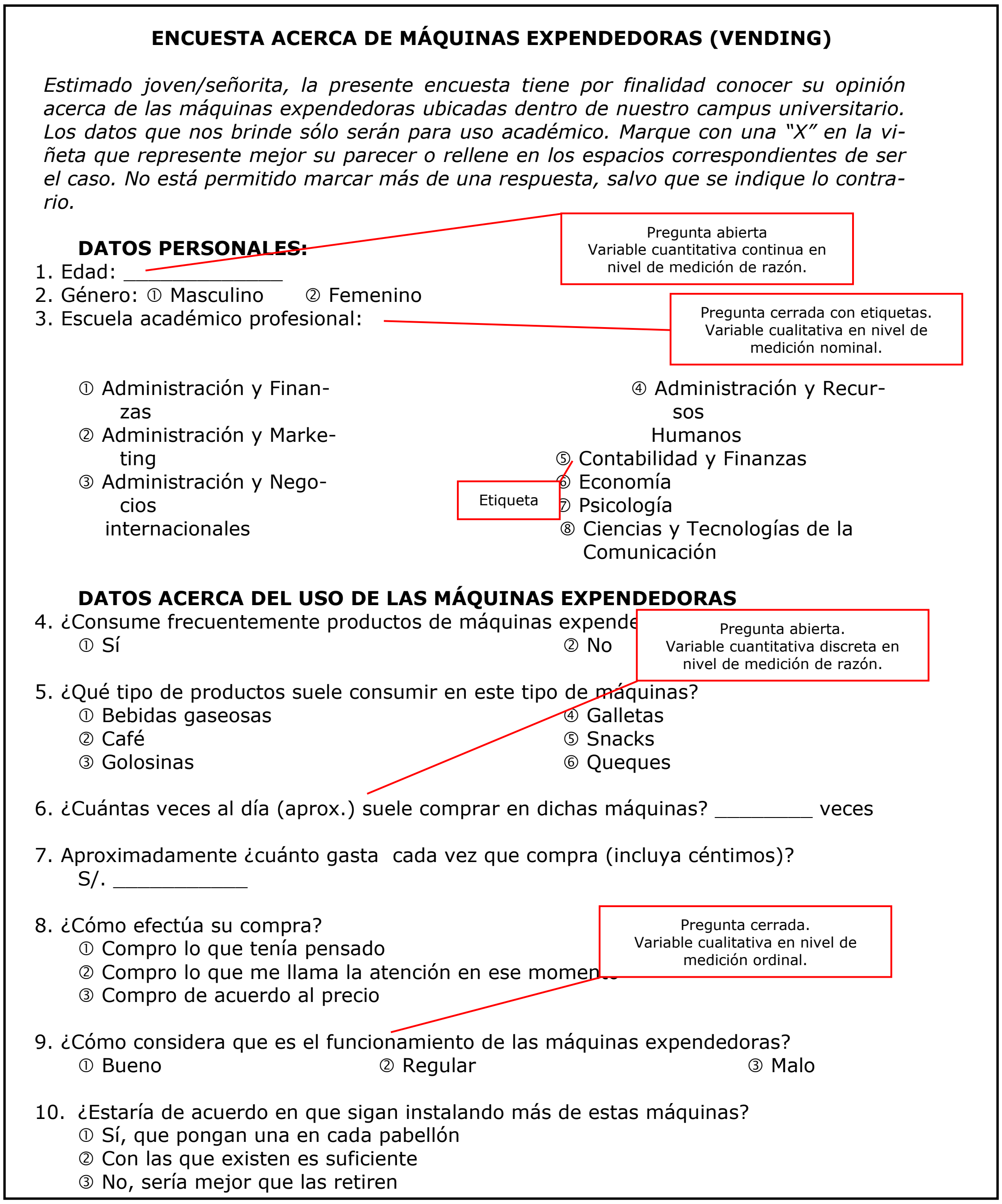
Un principio *extremadamente* importante en este capítulo, en este libro, y en la estadística en general, es que el método que se utiliza para reunir datos de muestras puede construir o destruir la validez de las conclusiones basadas en los datos. En la actualidad, a todos nosotros se nos bombardea con encuestas y resultados de encuestas. Algunas reúnen datos de muestras que son útiles porque describen de manera exacta características importantes de poblaciones. Otras encuestas usan datos muestrales recolectados de tal forma que condenan los resultados a la creciente pila de basura de la mala información.

ACTIVIDAD FORMATIVA Nº 1

Elabora una encuesta de 10 preguntas sobre un tema de tu carrera. En esta encuesta deberás considerar diferentes tipos de preguntas (abiertas y cerradas), diferentes tipos de variables (cualitativas, cuantitativas discretas y continuas), diferentes niveles de medición (nominal, ordinal, intervalo y razón). Las preguntas cerradas deberán incluir etiquetas numéricas.

Ejemplo:

ENCUESTA ACERCA DE MÁQUINAS EXPENDEDORAS (VENDING)



ENCUESTA ACERCA DE MÁQUINAS EXPENDEDORAS (VENDING)

Estimado joven/señorita, la presente encuesta tiene por finalidad conocer su opinión acerca de las máquinas expendedoras ubicadas dentro de nuestro campus universitario. Los datos que nos brinde sólo serán para uso académico. Marque con una "X" en la viñeta que represente mejor su parecer o rellene en los espacios correspondientes de ser el caso. No está permitido marcar más de una respuesta, salvo que se indique lo contrario.

DATOS PERSONALES:

Pregunta abierta
Variable cuantitativa continua en nivel de medición de razón.

1. Edad: _______________
2. Género: ① Masculino ② Femenino
3. Escuela académico profesional:

Pregunta cerrada con etiquetas.
Variable cualitativa en nivel de medición nominal.

① Administración y Finanzas
② Administración y Marketing
③ Administración y Negocios internacionales
④ Administración y Recursos Humanos
⑤ Contabilidad y Finanzas
⑥ Economía
⑦ Psicología
⑧ Ciencias y Tecnologías de la Comunicación

Etiqueta

DATOS ACERCA DEL USO DE LAS MÁQUINAS EXPENDEDORAS

4. ¿Consume frecuentemente productos de máquinas expende
① Sí ② No

Pregunta abierta.
Variable cuantitativa discreta en nivel de medición de razón.

5. ¿Qué tipo de productos suele consumir en este tipo de máquinas?
① Bebidas gaseosas
② Café
③ Golosinas
④ Galletas
⑤ Snacks
⑥ Queques

6. ¿Cuántas veces al día (aprox.) suele comprar en dichas máquinas? _________ veces

7. Aproximadamente ¿cuánto gasta cada vez que compra (incluya céntimos)?
S/. ____________

8. ¿Cómo efectúa su compra?
① Compro lo que tenía pensado
② Compro lo que me llama la atención en ese momento
③ Compro de acuerdo al precio

Pregunta cerrada.
Variable cualitativa en nivel de medición ordinal.

9. ¿Cómo considera que es el funcionamiento de las máquinas expendedoras?
① Bueno ② Regular ③ Malo

10. ¿Estaría de acuerdo en que sigan instalando más de estas máquinas?
① Sí, que pongan una en cada pabellón
② Con las que existen es suficiente
③ No, sería mejor que las retiren

GRACIAS POR PARTICIPAR EN ESTA ENCUESTA

INSTRUCCIONES

1. Estudia detenidamente los contenidos del Tema N° 1.
2. Elabora un breve marco teórico (2 hojas) respecto al tema del que trata tu encuesta. Con esta información estarás más enterado acerca del tema que elegiste, con lo cual tendrás mejores ideas para plantear su propia encuesta. Puedes fijarte algunos modelos de encuesta similares en el internet.
3. Contesta las siguientes preguntas acerca de tu encuesta:

- ¿Cuál es la población en estudio?
- ¿Cuál podría ser la muestra?
- ¿Cuáles son las unidades estadísticas?
- ¿Con los datos que has recogido puedes calcular parámetros o estadígrafos? ¿Por qué?

4. Elabora un pequeño comentario (5 líneas) acerca de por qué tendrías que tomar una muestra en vez de estudiar a toda la población.

Videos:

Tema Nº 1

Etapas de una investigación estadística

https://www.youtube.com/watch?v=fRvL6WGEF9U

Tema Nº 2

¿Cómo hacer una encuesta?

https://www.youtube.com/watch?v=q3Rgwxx3iI4

TEMA Nº 2:
TIPOS DE DISTRIBUCIONES Y GRÁFICA DE DATOS

Apreciado estudiante, ahora nos corresponde poner en práctica lo aprendido en el tema anterior para efectuar la organización de datos y luego su representación gráfica. La importancia de este tema radica en que aprenderás a organizar, resumir y representar los datos de una forma conveniente y comprensible. La representación que elijamos dependerá del tipo de variable y de dato con que estemos trabajando. El objetivo principal es analizar los datos y entender lo que indican.

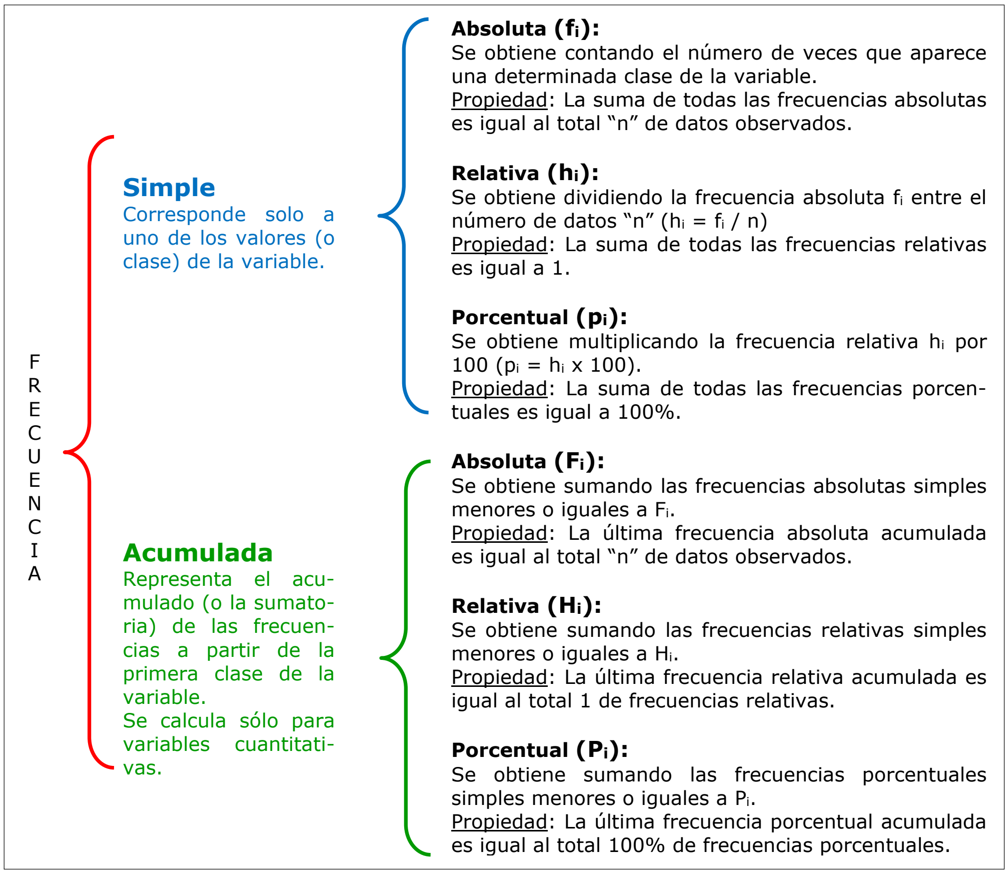
1. DISTRIBUCIÓN DE FRECUENCIAS

Como menciona Martínez Bencardino (2012, p. 44), "La distribución de frecuencias es un método utilizado para organizar y resumir información. Bajo este método, los datos recolectados se ordenan y clasifican, indicándonos la frecuencia o número de veces que se repiten".

1.1. Definiciones básicas:

Clase: Es una división de la variable. Se denota como subíndice con la letra "i" y el número total de clases con "m".

Frecuencia: Viene a ser el número de veces que aparece o se repite una clase de la variable. Puede ser:



FRECUENCIA

Simple
Corresponde solo a uno de los valores (o clase) de la variable.

Absoluta (f_i):
Se obtiene contando el número de veces que aparece una determinada clase de la variable.
Propiedad: La suma de todas las frecuencias absolutas es igual al total "n" de datos observados.

Relativa (h_i):
Se obtiene dividiendo la frecuencia absoluta f_i entre el número de datos "n" ($h_i = f_i / n$)
Propiedad: La suma de todas las frecuencias relativas es igual a 1.

Porcentual (p_i):
Se obtiene multiplicando la frecuencia relativa h_i por 100 ($p_i = h_i$ x 100).
Propiedad: La suma de todas las frecuencias porcentuales es igual a 100%.

Acumulada
Representa el acumulado (o la sumatoria) de las frecuencias a partir de la primera clase de la variable.
Se calcula sólo para variables cuantitativas.

Absoluta (F_i):
Se obtiene sumando las frecuencias absolutas simples menores o iguales a F_i.
Propiedad: La última frecuencia absoluta acumulada es igual al total "n" de datos observados.

Relativa (H_i):
Se obtiene sumando las frecuencias relativas simples menores o iguales a H_i.
Propiedad: La última frecuencia relativa acumulada es igual al total 1 de frecuencias relativas.

Porcentual (P_i):
Se obtiene sumando las frecuencias porcentuales simples menores o iguales a P_i.
Propiedad: La última frecuencia porcentual acumulada es igual al total 100% de frecuencias porcentuales.

Figura 5. Tipos de frecuencias
Fuente: Elaboración propia

1.2. Distribuciones de frecuencias unidimensionales (1 variable)

Para variable cualitativa:

FRECUENCIAS SIMPLES

CLASES

Variable x_i	Frecuencia absoluta f_i	Frecuencia relativa h_i	Frecuencia porcentual p_i
x_1	f_1	$h_1 = f_1 / n$	$p_1 = h_1 . 100$
x_2	f_2	$h_2 = f_2 / n$	$p_2 = h_2 . 100$
⋮	⋮	⋮	⋮
x_m	f_m	$h_m = f_m / n$	$p_m = h_m . 100$
TOTAL	**n**	**1**	**100%**

Para variable cuantitativa discreta:

FRECUENCIAS SIMPLES FRECUENCIAS ACUMULADAS

CLASES

Variable x_i	Frecuencia absoluta f_i	Frecuencia relativa h_i	Frecuencia Porcentual p_i	Frecuencia absoluta acumulada F_i	Frecuencia relativa acumulada H_i	Frecuencia porcentual acumulada P_i
x_1	f_1	h_1	p_1	f_1	h_1	p_1
x_2	f_2	h_2	p_2	f_1+f_2	h_1+h_2	p_1+p_2
x_3	f_3	h_3	p_3	$f_1+f_2+f_3$	$h_1+h_2+h_3$	$p_1+p_2+p_3$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_m	f_m	h_m	p_m	n	1	100%
TOTAL	**n**	**1**	**100%**			

Para variable cuantitativa continua:

Sobre este tema Jorge Toma (2012) señala:

Cuando se utilizan intervalos de igual amplitud, el proceso de generación de tales intervalos es el siguiente:

a) En primer lugar, debe establecerse el número de intervalos que se va a utilizar. Es recomendable que dicho número esté entre 5 y 15. No existe una regla fija para determinar el número óptimo de intervalos. El criterio del investigador cumple un papel importante en la determinación del mismo. Es conveniente tener presente que con pocos intervalos no se mostrarán detalles significativos para el análisis; por otro lado, con un número elevado de intervalos aparecerán demasiados detalles que difícilmente seremos capaces de captar.

Como referencia, se puede utilizar la regla de Sturges, la cual indica que el número de intervalos es dado por:

$$\mathbf{m = 1 + 3.3 \log_{10} (n)}$$

Donde n = Número de observaciones disponibles. El valor de "m" debe ser redondeado al entero más cercano.

b) Determinar el valor del rango o amplitud (R) de los datos, es decir, determinar:

$$\mathbf{R = x_{máx} - x_{mín}}$$

Donde: $x_{máx}$ = Observación mayor ; $x_{mín}$ = Observación menor

c) Determinar el tamaño o amplitud de cada intervalo (A):

$$\mathbf{A = R/m}$$

Cuando el cociente R/m no es exacto, el valor de A debe ser redondeado al valor superior más cercano, según las cifras decimales de los datos.

d) Generar los límites de los intervalos. Para el primer intervalo se considera como límite inferior al valor de la observación de menor magnitud, es decir, $L_i = x_{mín}$. El límite superior se obtiene sumando la amplitud A al $L_i = x_{mín}$.

Los demás intervalos se obtienen sumando la amplitud al límite superior del intervalo anterior.

e) Cada uno de los intervalos se considera cerrado a la izquierda y abierto a la derecha ($[L_i - L_s >$), es decir, se considera desde L_i a menos de L_s.

Esta regla no se aplica necesariamente para el último intervalo. Si el último límite superior tiene el mismo valor que la observación de mayor magnitud, deberá considerarse cerrado en ambos extremos ($[L_i - L_s]$), es decir, se considera desde L_i hasta L_s.

f) Cuando se realiza la clasificación de los datos y se generan las frecuencias absolutas, se pierde la identidad de los datos y, al final del proceso, solamente nos queda claro cuántos datos hay dentro de cada intervalo, y ya no es posible identificar sus valores. Por esta razón, cuando se desea calcular una medida estadística a partir de datos clasificados en intervalos de clase, se trabaja con los valores representativos de cada intervalo. A estos valores que representan la información de las observaciones de cada intervalo se les llama "marcas de clase" y, numéricamente, se obtienen como la semisuma de los límites inferior y superior de cada intervalo. Es decir:

$$y_i = (L_i + L_s) / 2$$

FRECUENCIAS SIMPLES **FRECUENCIAS ACUMULADAS**

INTERVALOS

Variable x_i	**Marca de clase y_i**	**Frec. Absol. f_i**	**Frec. Relat. h_i**	**Frec. Porc. p_i**	**Frec. Absol. Acum. F_i**	**Frec. Relat. Acum. H_i**	**Frec. Porc. Acum. P_i**
$[L_{i1}-L_{s1}>$	$\frac{L_{i1}+L_{s1}}{2}$	f_1	h_1	p_1	f_1	h_1	p_1
$[L_{s1}-L_{s2}>$	$\frac{L_{s1}+L_{s2}}{2}$	f_2	h_2	p_2	f_1+f_2	h_1+h_2	p_1+p_2
$[L_{s2}-L_{s3}>$	$\frac{L_{s2}+L_{s3}}{2}$	f_3	h_3	p_3	$f_1+f_2+f_3$	$h_1+h_2+h_3$	$p_1+p_2+p_3$
. . .		. . .	. . .	. . .	. . .	. . .	. . .
$[L_{sm-1}-L_{sm}>$	$\frac{L_{sm-1}+L_{sm}}{2}$	f_m	h_m	p_m	n	1	100%
TOTAL		**n**	**1**	**100%**			

1.3. Distribuciones de frecuencias bidimensionales (2 variables)

Llamadas también bivariadas, denotadas por (x_i ; y_i) y son usadas cuando medimos u observamos simultáneamente dos variables en cada uno de los individuos en estudio.

Y_i / **X_i**	y_1	y_2	y_3		y_i	**TOTAL (horizontal)**
x_1	f_{11}	f_{12}	f_{13}	. . .	f_{1i}	$f_{1.}$
x_2	f_{21}	f_{22}	f_{23}	. . .	f_{2i}	$f_{2.}$
x_3	f_{31}	f_{32}	f_{33}	. . .	f_{3i}	$f_{3.}$
. . .	. . .	. . .	. . .	. . .	. . .	. .
x_k	f_{k1}	f_{k2}	f_{k3}	. . .	f_{ki}	$f_{k.}$
TOTAL (vertical)	$f_{.1}$	$f_{.2}$	$f_{.3}$	. . .	$f_{.i}$	n

FRECUENCIAS CONJUNTAS

FRECUENCIAS MARGINALES DE LA VARIABLE "X"

FRECUENCIAS MARGINALES DE LA VARIABLE "Y"

Donde: $\mathbf{n = f_{.1}+f_{.2}+f_{.3}+ . . . +f_{.i} = f_{1.}+f_{2.}+f_{3.}+ . . . +f_{k.}}$

2. GRÁFICOS ESTADÍSTICOS:

Como menciona el INEI (2009, p. 13), "Gráfico estadístico: Llamado también "diagrama" es una representación visual de datos estadísticos por medio de puntos, líneas, barras, polígonos o figuras asociadas a escalas de medición, que permite una fácil comprensión de la información en su conjunto".

Todo gráfico estadístico debe tener un código o número, título y cuerpo. Así, los elementos de un gráfico estadístico son:

Código o número de gráfico:

Elemento numérico que permite identificar al gráfico estadístico. Si se presenta más de un gráfico en un capítulo, cada uno debe incluir el número del capítulo seguido de un punto y el número de gráfico correspondiente. Ejemplo: *Figura 5.4*

Título:

Se coloca después del número de gráfico, con el propósito de dar a conocer las variables y las características contenidas en él. El título expresará el contenido del gráfico en forma ordenada, clara y breve, evitando la descripción excesiva o la brevedad extrema en la descripción del contenido de la información. El título debe estar redactado en mayúsculas y con tildes.

Debe responder a 4 preguntas básicas: a) ¿Dónde? Se refiere al lugar al que corresponde la información. b) ¿Qué? Se refiere al hecho observado o característica principal que se quiere mostrar. c) ¿Cómo? Se refiere a cómo se presenta la información, empezando por la leyenda del gráfico, que irán precedidas por la preposición "POR" y continuando con el eje de conceptos, que irán precedidas por la preposición "SEGÚN". d) ¿Cuándo? Se refiere al período temporal que cubre la información. Puede estar referido a una semana, mes, trimestre, año o, a un día determinado.

Ejemplo 1: *PUNO: TASA DE ASISTENCIA ESCOLAR, POR SEXO, SEGÚN ÁREA DE RESIDENCIA, 2007*

¿Dónde? : Puno

¿Qué? : Tasa de asistencia escolar

¿Cómo? : Por sexo, según área de residencia

¿Cuándo? : 2007

Ejemplo 2: *PERÚ: PRODUCCIÓN DE GAS NATURAL SEGÚN ZONAS GEOGRÁFICAS, 2001-08*

(Miles de barriles)

¿Dónde? : Perú

¿Qué? : Producción de gas natural

¿Cómo? : Según zonas geográficas

¿Cuándo? : Del 2001 al 2008

Cuerpo:

Es la ilustración de los valores asociados a los datos presentados mediante los siguientes elementos

- **Figura:** Conjunto de puntos, líneas, barras, polígonos o figuras utilizados en la representación de los datos estadísticos.
- **Escala o eje de valores:** Es la línea vertical recta segmentada que representa la escala de medición a la que corresponden los datos estadísticos del gráfico. Si se visualizan los rótulos de datos, es optativo mostrar u ocultar el eje de valores.

- **Leyenda:** Es la descripción de la simbología utilizada, sea ésta mediante colores, densidades de color, sombreados o tipos de línea usados para diferenciar conceptos a los que se refieren los datos estadísticos. Corresponde a las características que en el título, generalmente, vienen precedidas de la preposición "por".

- **Eje de conceptos:** Es la expresión específica de cada uno de los conceptos y valores a los que se refieren los datos. Corresponde a las características que en el título, generalmente, vienen precedidas de la preposición "según".

Pie:

Constituye la parte inferior del gráfico y comprende las notas, llamadas y fuente. Se destina para anotar aquellas aclaraciones o señalamientos particulares y generales sobre la información, necesarios para una mejor interpretación de ésta por parte del usuario. Si los gráficos están acompañados de cuadros, el uso del pie de gráfico es optativo.

- **Nota:** Es la información de carácter general sobre el contenido del gráfico (definiciones). Se usa también para indicar la metodología adoptada en la investigación o elaboración de los datos.

- **Llamada:** Es la información específica aplicable a determinada parte del cuerpo del gráfico que se utiliza con el objeto de hacer aclaraciones particulares sobre la interpretación conceptual, cobertura geográfica o referencia temporal de los datos estadísticos ofrecidos.

- **Fuente:** Es la indicación que se encuentra al pie del gráfico y tiene como fin cumplir un triple propósito: Otorgar el crédito correspondiente a la entidad responsable de producir la información, señalar la unidad o departamento que genera la información y el documento de donde se obtiene, así como orientar al usuario sobre su localización en caso de requerir alguna consulta directa. Si los gráficos están acompañados de cuadros que tienen el mismo origen de los datos, el uso de la "fuente" es optativo.

La fuente debe escribirse seguida de dos puntos y alineada al margen izquierdo del gráfico. Se ubicará después de la nota y las llamadas, o en lugar de ellas cuando alguna o ambas no existan.

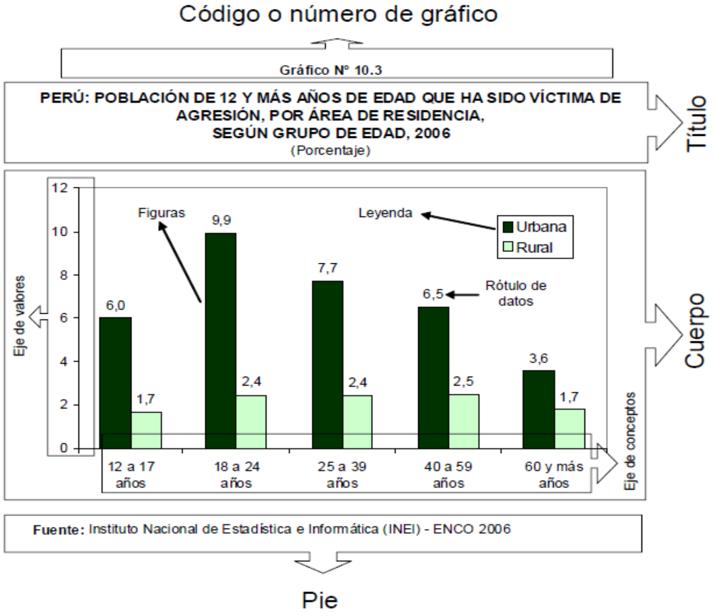


Figura 6. Tipos de frecuencias
Fuente: Instituto Nacional de Estadística de Informática – INEI

2.1. PRINCIPALES TIPOS DE GRÁFICOS:

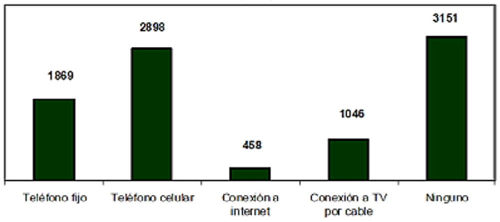
Gráfico de barras simples: Compara valores entre categorías de una variable.

Barras verticales

Gráfico Nº 5.10

PERÚ: HOGARES EN VIVIENDAS PARTICULARES, SEGÚN SERVICIO DE INFORMACIÓN Y COMUNICACIÓN QUE DISPONE EL HOGAR, 2007

(Miles)



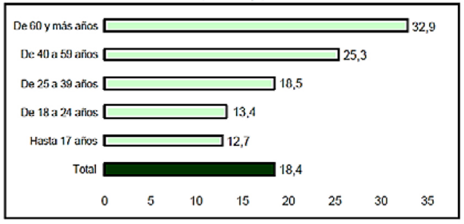
Fuente: INEI - Censos Nacionales 2007: XI de Población y VI de Vivienda.

Barras horizontales

Gráfico Nº 8.6

PERÚ: POBLACIÓN QUE APRENDIÓ QUECHUA, AYMARA U OTRA LENGUA NATIVA EN SU NIÑEZ, SEGÚN GRUPO DE EDAD, 2006

(Porcentaje)



Fuente: INEI - Encuesta Nacional Continua 2006.

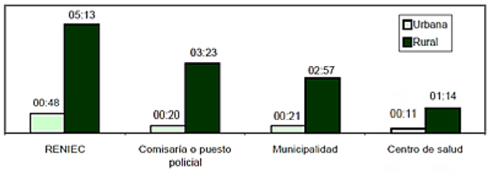
Gráfico de barras comparativas, de dos o más series. Se utilizan para comparar la magnitud de dos o más variables mediante barras que pueden colocarse juntas.

Barras verticales

Gráfico Nº 12.5

PERÚ: TIEMPO PROMEDIO DE DESPLAZAMIENTO A UN ESTABLECIMIENTO PÚBLICO, POR ÁREA DE RESIDENCIA, ENE-JUN 2006

(Horas : minutos)



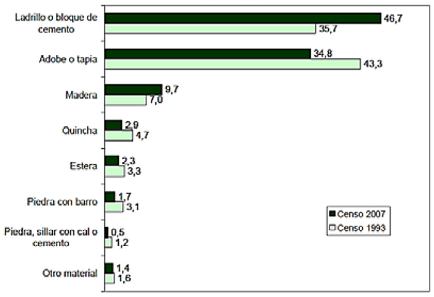
Fuente: INEI - Encuesta Nacional Continua 2006.

Barras horizontales

Gráfico Nº 9.9

PERÚ: VIVIENDAS PARTICULARES, SEGÚN MATERIAL PREDOMINANTE EN LAS PAREDES EXTERIORES, 1993 Y 2007

(Porcentaje)



Fuente: INEI - Censos Nacionales de Población y Vivienda 1993 y 2007.

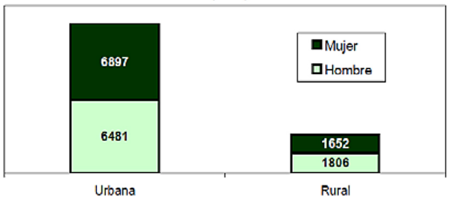
Barras Apiladas: Compara entre categorías el aporte de cada valor en el total.

Barras verticales

Gráfico Nº 8.25

PERÚ: POBLACIÓN CENSADA DE 18 Y MÁS AÑOS DE EDAD QUE TIENE DOCUMENTO NACIONAL DE IDENTIDAD, POR SEXO, SEGÚN ÁREA DE RESIDENCIA, 2007

(Miles)



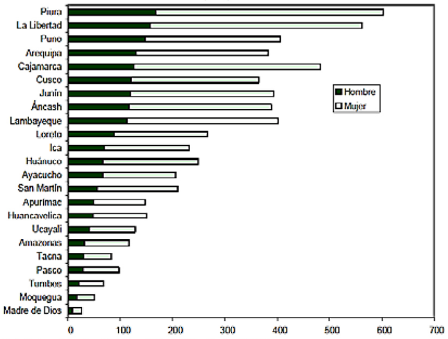
Fuente: INEI - Censos Nacionales 2007: XI de Población y VI de Vivienda.

Barras horizontales

Gráfico Nº 2.17

PERÚ: POBLACIÓN ECONÓMICAMENTE INACTIVA DE 14 Y MÁS AÑOS DE EDAD, POR SEXO, SEGÚN DEPARTAMENTO, 2007

(Miles)



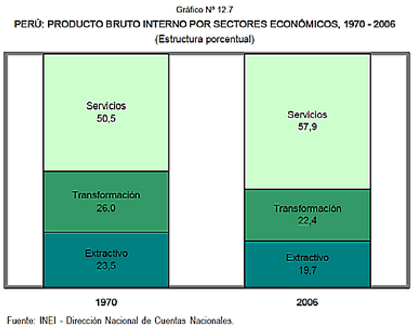
Nota: Excluye el departamento de Lima y la Provincia Constitucional del Callao.

Fuente: INEI - Censos Nacionales 2007: XI de Población y VI de Vivienda.

Fuente: Instituto Nacional de Estadística de Informática – INEI

Gráfico rectangular: Es un diagrama compuesto por un rectángulo dividido en porciones, de tal manera que cada una de estas porciones es proporcional a la magnitud del valor que representa.

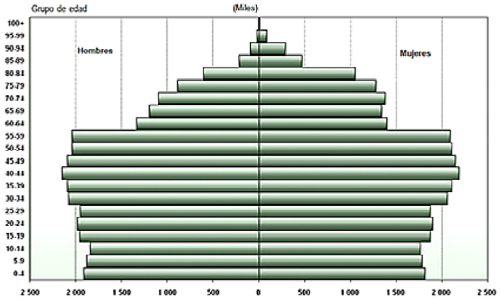
Gráfico Nº 12.7
PERÚ: PRODUCTO BRUTO INTERNO POR SECTORES ECONÓMICOS, 1970 - 2006
(Estructura porcentual)



Fuente: INEI - Dirección Nacional de Cuentas Nacionales.

Pirámide de población: La pirámide de población es un tipo de gráfico de eje central, donde el conjunto estudiado es la población de un lugar dividida en grupos de edad y sexo.

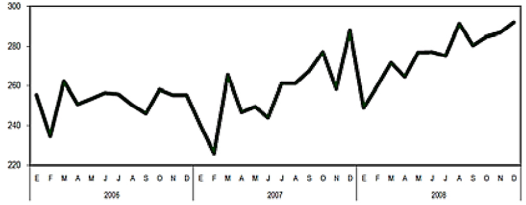
Gráfico 8.3
FRANCIA: PIRÁMIDE DE POBLACIÓN, 2005
(Miles)



Fuente: Organización de las Naciones Unidas (ONU) - Proyecciones de la población mundial, Revisión 2004.

Gráfico de línea simple: Muestra las variaciones de un fenómeno a través de un determinado período.

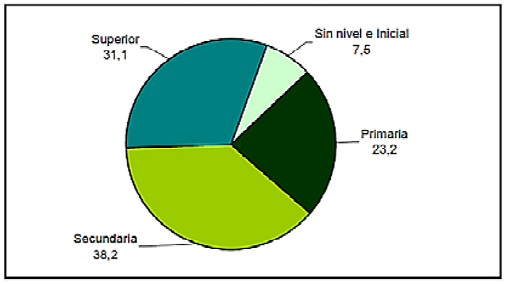
Gráfico Nº 3.2
PERÚ: INDICE DE PRODUCCIÓN MINERO METÁLICA: 2006 - 2008
(Año base 1994=100)



Fuente: INEI - Dirección Técnica de Indicadores Económicos.

Gráfico circular o de torta: Permite ver la distribución interna de los datos que representan un hecho, en forma de porcentajes sobre un total.

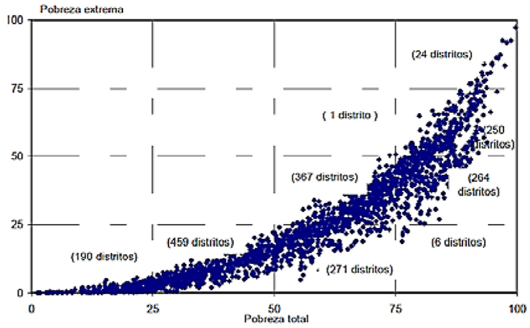
Gráfico Nº 5.2
PERÚ: POBLACIÓN CENSADA DE 15 Y MÁS AÑOS DE EDAD, SEGÚN NIVEL EDUCATIVO, 2007
(Distribución porcentual)



Fuente: INEI - Censos Nacionales 2007: XI de Población y VI de Vivienda.

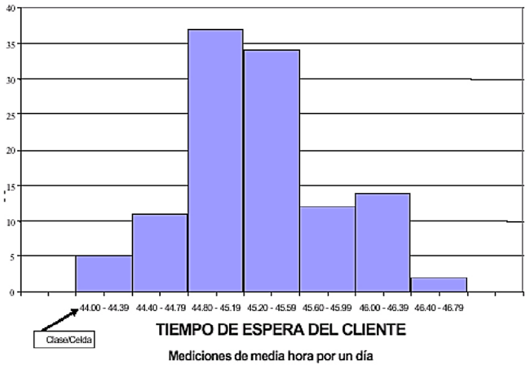
Diagrama de dispersión: Muestra la asociación o relación que existe entre dos variables unidas a un mismo fenómeno.

Gráfico Nº 3.2
PERÚ: DISTRIBUCIÓN DE LOS DISTRITOS POR POBREZA TOTAL Y POBREZA EXTREMA, 2007



Fuente: INEI - Censos Nacionales 2007: XI de Población y VI de Vivienda.

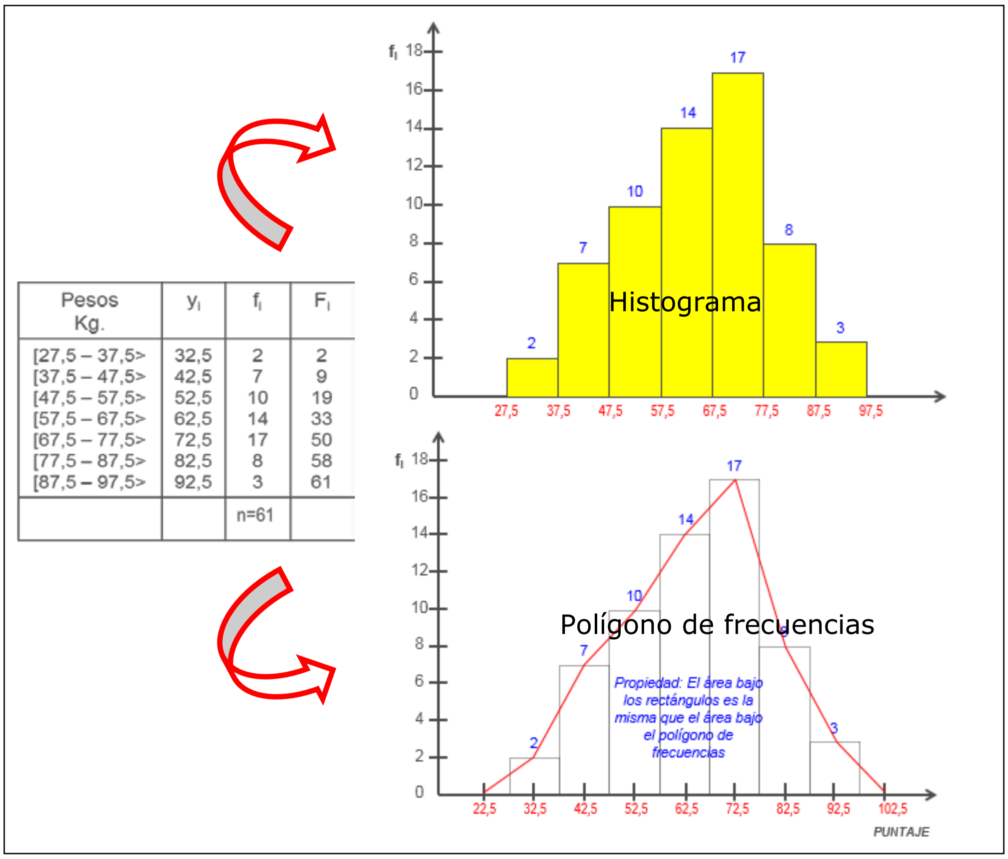
Histograma de Frecuencias: Se utiliza para mostrar la distribución de frecuencias de variables cuantitativas. Se construye dibujando barras contiguas que tienen como base la amplitud de cada intervalo y como alturas las frecuencias respectivas.



Fuente: Instituto Nacional de Estadística de Informática–INEI

2.2. HISTOGRAMA, POLÍGONO DE FRECUENCIAS Y OJIVA "MENOR QUE":

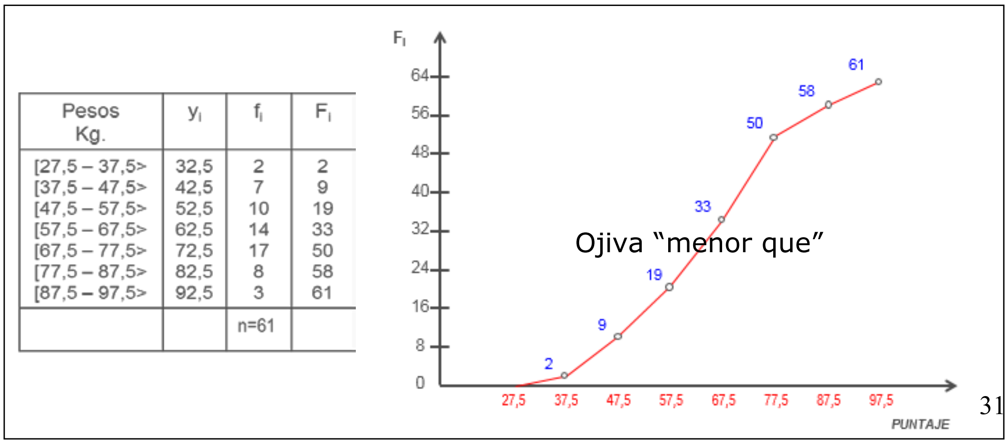
Lind et al. (2001) señalan que **el histograma** es un gráfico en el que las clases se señalan en el eje horizontal y las frecuencias de clase en el eje vertical. Las frecuencias de clase se representan por medio de las alturas de las barras, que se dibujan de manera adyacente. **El polígono de frecuencias** al igual que el histograma también muestra la forma que tiene una distribución y es similar al histograma, pero se elabora a partir de las marcas de clase. Al igual que el histograma se elabora con cualquier frecuencia simple.



Pesos Kg.	y_i	f_i	F_i
[27,5 – 37,5>	32,5	2	2
[37,5 – 47,5>	42,5	7	9
[47,5 – 57,5>	52,5	10	19
[57,5 – 67,5>	62,5	14	33
[67,5 – 77,5>	72,5	17	50
[77,5 – 87,5>	82,5	8	58
[87,5 – 97,5>	92,5	3	61
		n=61	

Figura 7. Histograma y polígono de frecuencia
Fuente: Elaboración propia

La ojiva se emplea para representar cualquiera de las frecuencias acumuladas. Es como un polígono de frecuencias, con la diferencia que se elabora con otro tipo de frecuencias.



Pesos Kg.	y_i	f_i	F_i
[27,5 – 37,5>	32,5	2	2
[37,5 – 47,5>	42,5	7	9
[47,5 – 57,5>	52,5	10	19
[57,5 – 67,5>	62,5	14	33
[67,5 – 77,5>	72,5	17	50
[77,5 – 87,5>	82,5	8	58
[87,5 – 97,5>	92,5	3	61
		n=61	

Figura 8. Ojiva menor
Fuente: Elaboración propia

2.3. EL DIAGRAMA DE PARETO:

Fue creado por Vilfredo Pareto, economista y sociólogo (1848 – 1923) quien publicó los resultados de sus estudios sobre la distribución de la riqueza, observando que el 80% de la misma se encontraba concentrada en el 20% de la población.

La relación 80/20 ha sido encontrada en distintos campos. Por ejemplo, el 80% de los problemas de una organización son debidos a un 20% de las causas posibles. Evidentemente, la relación no debe ser exactamente 80/20, pero sí se puede aventurar que unas pocas causas son responsables de la mayor parte de los problemas.

Una importante aplicación del Principio de Pareto está en el diseño de programas de mejora de la calidad.

PRINCIPIO DE PARETO

En todo grupo de elementos o factores que contribuyen a un mismo efecto, unos pocos son responsables de la mayor parte de dicho efecto.

ANÁLISIS DE PARETO

El objetivo de esta comparación es clasificar dichos elementos o factores en dos categorías: Las "Pocas Vitales" (los elementos muy importantes en su contribución) y los "Muchos Triviales" (los elementos poco importantes en ella).

Ejemplo: Los clientes de un banco han venido presentando sus quejas a la Gerencia respecto a los siguientes motivos:

MOTIVO DE QUEJA	FRECUENCIA (FI)
Demora en la atención	18
Dificultad en el parqueo	22
Horario de atención inadecuado	43
Falta de información general	28
Local no apropiado	9

Elabore el diagrama de Pareto e identifique los aspectos que se deben de atender prioritariamente para que los clientes no se quejen.

Solución:

- Primero: Organizar las quejas en una tabla de frecuencias incluyendo sus porcentajes.
- Segundo: Ordenar las categorías en forma decreciente, en función a sus frecuencias o porcentajes, incluyendo los porcentajes acumulados.

MOTIVO DE QUEJA	FRECUENCIA (FI)	PORCENTAJE (PI%)	PORCENTAJE ACUMUL. (PI%)
Horario de atención inadecuado	43	36%	36%
Falta de información general	28	23%	59%
Dificultad en el parqueo	22	18%	77%
Demora en la atención	18	15%	92%
Local no apropiado	9	8%	100%
	n = 120	100%	

- Tercero: Trazar y rotular el eje horizontal y a ambos lados trazar dos ejes verticales (a la izquierda van las frecuencias absolutas, a la derecha las frecuencias porcentuales acumuladas).

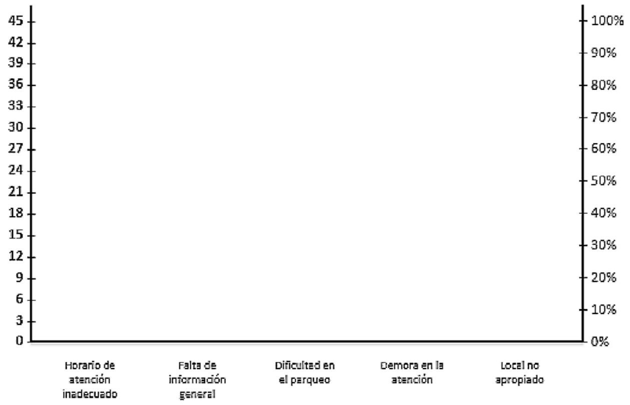


Figura 9. Solución
Fuente: Elaboración propia

- Cuarto: Dibujar un gráfico de barras que muestre el efecto de cada uno de los elementos contribuyentes.

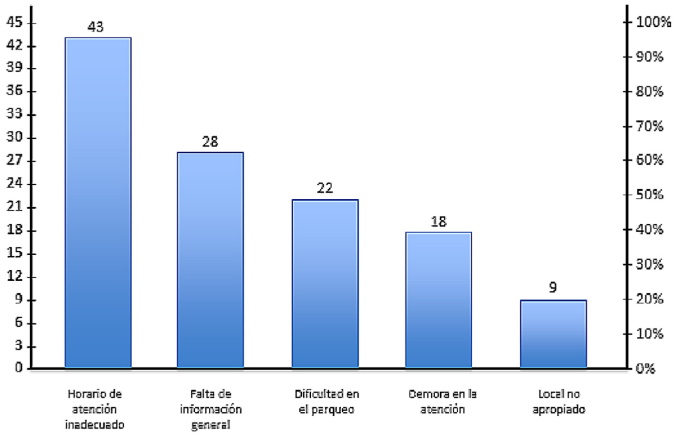


Figura 10. Solución
Fuente: Elaboración propia

- Quinto: Trazar un gráfico lineal cuyos puntos representan el porcentaje acumulado de la tabla.

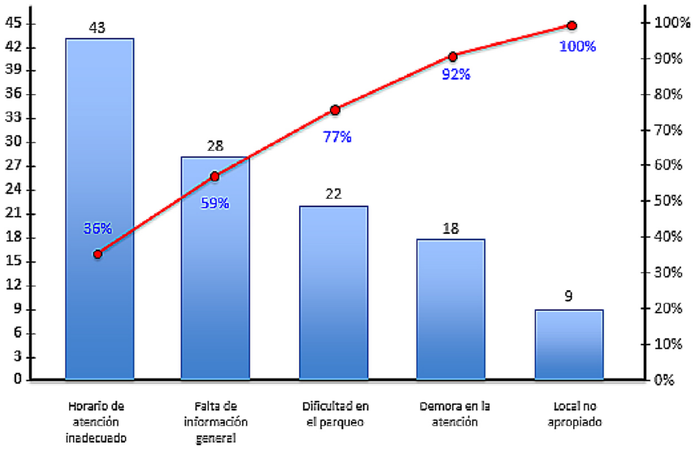


Figura 11. Solución
Fuente: Elaboración propia

- Sexto: Diferenciar los elementos vitales y triviales trazando una línea horizontal a partir el eje "y" a la altura del 80% y de ahí hacia el eje "x". Las categorías de la izquierda representan los elementos vitales y los de la derecha los triviales.

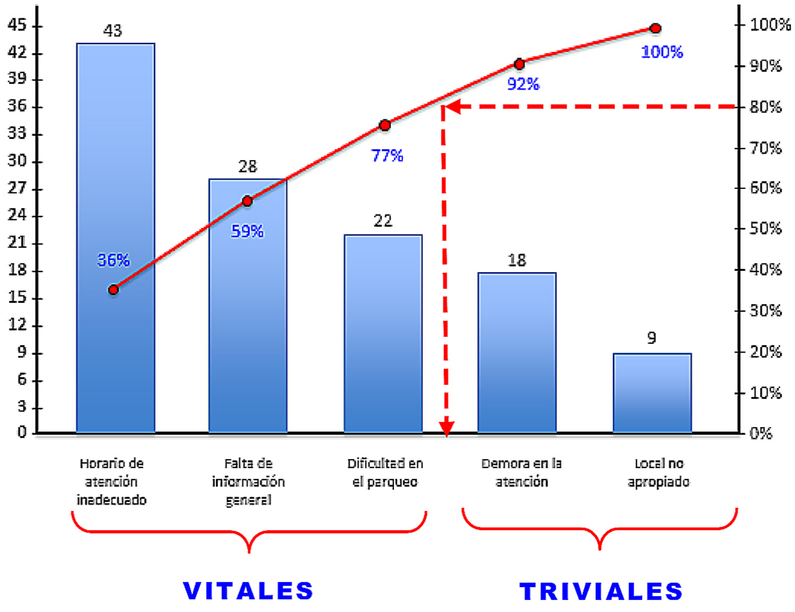


Figura 12. Diagrama de Pareto
Fuente: Elaboración propia

De acuerdo al diagrama de Pareto, el Gerente debe de dar solución a las quejas relacionadas al horario de atención, falta de información y dificultad en el parqueo. De esa manera estará solucionando el 80% de las quejas presentadas por los clientes.

LECTURA SELECCIONADA N° 2:

Leer apartado: ¿Los Premios de la Academia discriminan por la edad?

Triola, M.(2009). *Estadística.* México: Pearson Educación.

Cada año se otorgan Óscares a la mejor actriz y al mejor actor. En la tabla se presenta una lista con las edades de los galardonados en el momento de la ceremonia de entrega de los premios. Las edades aparecen en orden, empezando con la primera ceremonia de los Premios de la Academia en 1928. [Notas: En 1968 hubo un empate en la categoría de mejor actriz, y se utilizó el promedio (la media) de las dos edades; en 1932 hubo un empate en la categoría de mejor actor, y se utilizó el promedio (la media) de las dos edades. Tales datos se basan en el artículo "Ages of Oscar-winning Best Actors and Actresses", de Richard Brown y Gretchen Davis, en la revista Mathematics Teacher. En ese artículo, el año de nacimiento del ganador del premio se restó del año de la ceremonia; no obstante, las edades de la tabla se basan en la fecha de nacimiento del ganador y en la fecha de la ceremonia de premiación]. La pregunta básica que consideraremos es: ¿Hay diferencias importantes entre las edades de las mejores actrices y las edades de los mejores actores? ¿Al parecer los actores y las actrices son juzgados estrictamente por sus habilidades artísticas? O hien, ¿existe discriminación por la edad y las mejores actrices suelen ser más jóvenes que los mejores actores? ¿Hay algunas otras diferencias evidentes? Además de ser interesante, esto es importante porque nos brinda información sobre la forma en que nuestra sociedad percibe a los hombres y a las mujeres en general.

Una comparación visual entre las edades de la tabla sería reveladora para las personas que tienen una habilidad especial para observar un orden en este tipo de listas de números; sin embargo, para nosotros los simples mortales, es probable que la lista no revele mucha información. Afortunadamente, se dispone de métodos para investigar este tipo de conjuntos de datos, y pronto veremos que tales procedimientos revelan características importantes que nos permiten entender los datos. Seremos capaces de hacer comparaciones inteligentes y reveladoras; aprenderemos técnicas para resumir, graficar, describir, explorar y comparar conjuntos de datos como los de la tabla.

Tabla
Premios de la Academia: Edades de las mejores actrices y los mejores actores
Las edades (en años) aparecen en orden, empezando con la primera ceremonia de premiación.

Mejores actrices	Mejores actores
22 37 28 63 32 26 31 27 27 28	44 41 62 52 41 34 34 52 41 37
30 26 29 24 38 25 29 41 30 35	38 34 32 40 43 56 41 39 49 57
35 33 29 38 54 24 25 46 41 28	41 38 42 52 51 35 30 39 41 44
40 39 29 27 31 38 29 25 35 60	49 35 47 31 47 37 57 42 45 42
43 35 34 34 27 37 42 41 36 32	44 62 43 42 48 49 56 38 60 30
41 33 31 74 33 50 38 61 21 41	40 42 36 76 39 53 45 36 62 43
26 80 42 29 33 35 45 49 39 34	51 32 42 54 52 37 38 32 45 60
26 25 33 35 35 28	46 40 36 47 29 43

El problema incluye las edades de los ganadores del Óscar a la mejor actriz y al mejor actor. Con lo aprendido utilizamos distribuciones de frecuencias y gráficas para investigar si las edades de las actrices eran significativamente diferentes de las edades de los actores. Con base en los resultados obtenidos, parece que las actrices ganadoras del Óscar son más jóvenes que los actores ganadores de este premio.

ACTIVIDAD FORMATIVA N° 2

Elabora las tablas de frecuencia y los gráficos estadísticos que se piden en cada uno de los ejercicios. Luego contesta las preguntas que se hacen.

INSTRUCCIONES

1. Según la Asociación de lucha contra la Bulimia y la Anorexia, las pautas culturales han determinado que la delgadez sea sinónimo de éxito social. Muchos jóvenes luchan para conseguir el "físico ideal" motivados por modelos, artistas o por la publicidad comercial. Durante el mes de marzo del año 2014, en la Universidad Agraria "La Molina" de la ciudad de Lima, después de las vacaciones de verano, se observó con precaución a 27 alumnos con síntomas de anorexia, registrándose los siguientes signos visibles: Dieta Severa, Miedo a Engordar, Hiperactividad, Uso de Ropa Holgada, Dieta Severa, Uso de Laxantes, Miedo a Engordar, Dieta Severa, Uso de Ropa Holgada, Dieta Severa, Uso de Ropa Holgada, Dieta Severa, Dieta Severa, Dieta Severa, Uso de Ropa Holgada, Hiperactividad, Uso de Laxantes, Miedo a Engordar, Uso de Laxantes, Dieta Severa, Uso de Ropa Holgada, Uso de Laxantes, Hiperactividad, Uso de Laxantes, Uso de Ropa Holgada, Hiperactividad, Dieta Severa.

   - Resume la información anterior en una tabla de distribución de frecuencias. Determina: Población, muestra, unidad estadística, variable, tipo de variable, nivel de medición.

- ¿Cuál de los síntomas es el que mayor presencia tiene en dicha muestra? Indique el porcentaje.

Elabora el gráfico correspondiente para la variable en estudio.

2. Se ha tomado nota del número de tardanzas que, en el transcurso del mes, han tenido un grupo de trabajadores administrativos en una oficina de la SUNAT – Huancayo, obteniéndose el siguiente resultado:

1	2	2	1	3	2	6	1	2	3
0	0	1	2	6	3	4	3	2	4
2	3	2	3	2	2	2	1	3	2
4	4	3	1	1	1	4	2	3	1
5	1	4	5	4	5	5	0	1	0

Determina la población, la muestra, la variable y tipo de variable.

- Elabora una tabla de distribución de frecuencias.
- Elabora el gráfico correspondiente para la variable en estudio.

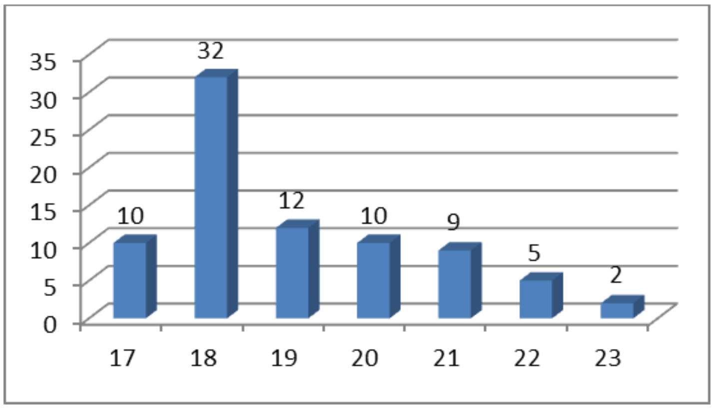
3. Los siguientes datos representan las estaturas de 60 mujeres hospitalizadas en la Maternidad de Lima en el mes de octubre del año 2014:

1,70	1,65	1,50	1,66	1,78	1,60	1,59	1,70	1,63	1,58	1,60	1,64
1,70	1,66	1,63	1,58	1,70	1,69	1,65	1,67	1,73	1,74	1,57	1,68
1,69	1,68	1,71	1,70	1,73	1,78	1,71	1,65	1,77	1,72	1,65	1,80
1,78	1,68	1,65	1,65	1,58	1,67	1,65	1,60	1,68	1,60	1,63	1,76
1,80	1,65	1,68	1,62	1,70	1,69	1,61	1,68	1,71	1,65	1,53	1,62

Ordena la información proporcionada en una tabla de frecuencias con intervalos de igual amplitud.

- ¿Cuál es la estatura que presentan la mayoría de mujeres? Determina* el porcentaje.
- Elabora el gráfico correspondiente para la variable en estudio.

4. El grafico muestra las edades de los estudiantes ingresantes a la carrera de INGENIERA CIVIL de la UCCI en agosto del año 2014.



- ¿Qué título le darías al gráfico? Recuerda que el título se redacta respondiendo a las preguntas: Dónde, qué, cómo, cuándo.
- Determina el valor de verdad (V) de las siguientes proposiciones:
  - ✓ El 40 % de los ingresantes tienen la edad de 18 años ( )
  - ✓ La muestra corresponde a los 80 estudiantes ingresantes de la carrera de Ingeniería Civil ( )
  - ✓ En un 90% de los estudiantes son mayores de edad ( )
  - ✓ Los estudiantes que tienen 18 años se diferencian de los estudiantes de 20 años en un 10% ( )

5. Los siguientes datos representan los precios (en miles de Dólares) de una muestra de 28 vehículos vendidos el año pasado en Automotores "San Jacinto" S.A. – Huancayo. (Los datos están ordenados de menor a mayor).

10,2	12,2	17,3	21,2	23,6	27,8	33,4
10,3	12,9	17,8	21,6	26,8	30,1	33,7
11,4	15,4	18,2	23,5	27,3	30,8	34,2
11,9	16,1	19,6	23,6	27,4	31,8	34,3

- Completa el siguiente cuadro:

• Población en estudio	
• Muestra (indique el tamaño)	
• Que variable está en estudio	
• Tipo de variable (cualitativa, cuantitativa discreta, cuantitativa continua)	
• Unidad estadística	
• Nivel de medición (Nominal, ordinal, intervalo, razón)	

- Organiza los datos haciendo uso de una tabla de distribución de frecuencias, agrupando en intervalos y extrae dos conclusiones que consideres relevantes.

6. Un cuestionario proporciona 58 respuestas "Sí", 42 "No" y 20 "sin opinión". En la elaboración de una gráfica circular, ¿cuántos grados mediría la sección del círculo que corresponde a las respuestas "Sí"?

   - ¿Cuántos grados mediría la sección del círculo que corresponde a las respuestas "No"?
   - Dibuja una gráfica circular para presentar dichos resultados.
   - Elabora una gráfica de barras para presentar dichos resultados.

7. El correo electrónico no solicitado y el spam afectan la productividad de los empleados de oficina. Una encuesta de Insight Express monitoreó a dichos empleados para determinar el tiempo improductivo por día dedicado a correo electrónico no solicitado y spam. Los datos siguientes presentan una muestra del tiempo en minutos dedicado a esta tarea.

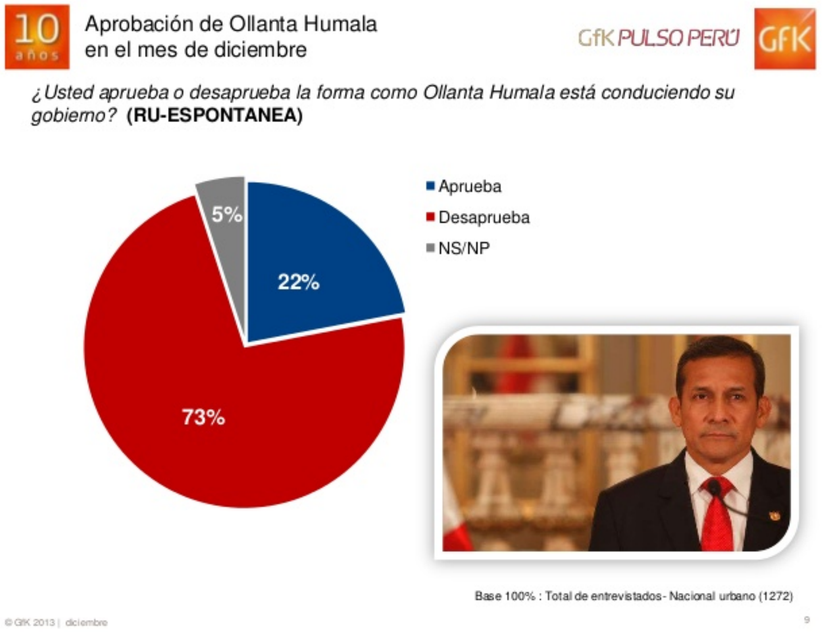
2	4	8	4
8	1	2	32
12	1	5	7
5	5	3	4
24	19	4	14

Resume los datos mediante la elaboración de lo siguiente:

- Una tabla de distribución de frecuencias en intervalos, incluyendo frecuencias simples y acumuladas.
- Elabora el histograma, polígono de frecuencias y ojiva para representar estos datos.

Responde: ¿Qué porcentaje de empleados de oficina pasó la mayor cantidad de tiempo dedica al correo electrónico no solicitado o spam? ¿Qué porcentaje pasó el menor tiempo al día en esta tarea?

8. Según encuesta elaborada por GfK PULSO PERÚ y aplicada en una muestra de 1272 personas en diciembre del 2013, la aprobación del Presidente Ollanta Humala era la siguiente:



10 años

Aprobación de Ollanta Humala en el mes de diciembre

GfK PULSO PERÚ GfK

¿Usted aprueba o desaprueba la forma como Ollanta Humala está conduciendo su gobierno? **(RU-ESPONTANEA)**

Base 100% : Total de entrevistados - Nacional urbano (1272)

© GfK 2013 | diciembre

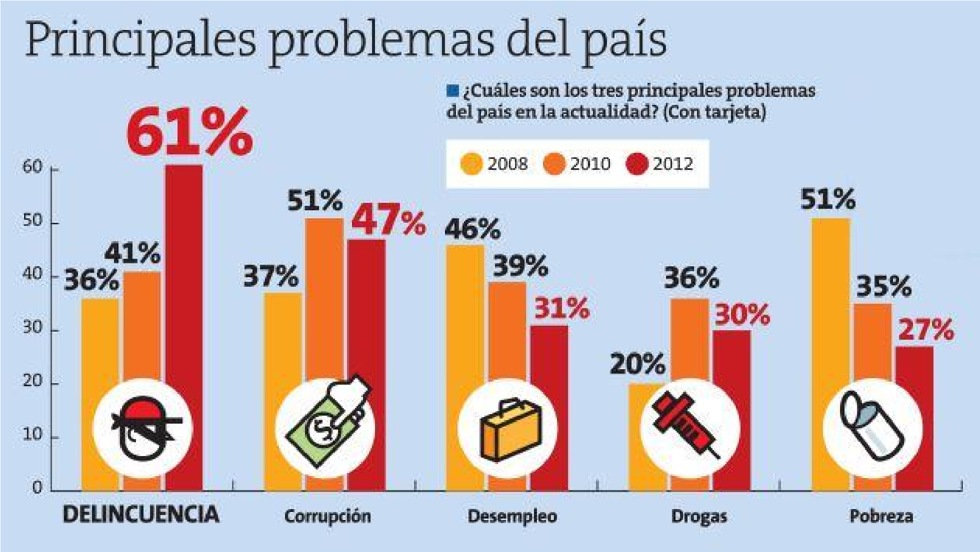
9

Responde las siguientes interrogantes:

- ¿Cuántas personas aprobaban la forma como Ollanta Humala conducía su gobierno?

- En la muestra de 1272 personas, ¿en cuánto se diferenciaban las personas que aprobaban y las que desaprobaban la forma en que conducía su gobierno el presidente de la república?
- Convierta el gráfico de sectores en uno de barras.

9. ENCUESTA PROÉTICA. A diferencia de años anteriores, la inseguridad ciudadana ha desplazado a la corrupción como principal problema de la ciudadanía, según revela el último sondeo de Ipsos Apoyo.



Responde las siguientes interrogantes:

- ¿Cuál fue el principal problema del país el año 2012?
- ¿Qué tendencia muestra la delincuencia durante el período 2008–2012: Aumenta, disminuye o se mantiene? ¿En qué porcentaje varió del año 2010 al año 2012?
- ¿En qué año el desempleo fue el menor problema?
- ¿Cuál fue el menor problema que tuvo el país el año 2010?

10. LA VOZ DEL PUEBLO… *La polémica propuesta para que los militares apoyen a la Policía Nacional en la lucha contra la criminalidad tuvo eco en la población. Según la última encuesta de Idice del Perú, el 85.3 % de entrevistados está de acuerdo con esta iniciativa, mientras un 9.2 % está en contra y un 5.5 % no responde. (SEGÚN ENCUESTA DE IDICE DEL PERÚ).*

- Con los datos elabora una tabla de frecuencias porcentuales.
- Elabora un gráfico de sectores que representa la información presentada.

PRUEBA DE DESARROLLO – UNIDAD I

1. Una agencia de viajes de alcance nacional ofrece tarifas especiales de ciertas travesías por la selva peruana para personas mayores de edad. El gerente de esta empresa desea información adicional acerca de las edades de las personas que participan en tales viajes. Una muestra al azar de clientes que fueron a una travesía el mes pasado indicó registró las siguientes edades:

18	41	50	53	58	61	63	71
26	43	51	54	58	61	63	71
34	44	52	54	59	62	63	77
36	45	52	56	60	62	65	83
38	50	53	58	60	62	66	84

- Organice los datos en una distribución de frecuencias utilizando intervalos de clase.
- ¿Entre qué edades tienden los datos a acumularse?
- ¿Qué porcentaje de los clientes tienen al menos 48 años?
- ¿Qué porcentaje de clientes tienen a lo más de 57 años?
- ¿Cuántos clientes tienen las máximas edades en la tabla?

2. Elabore el histograma de las edades del ejercicio anterior y trace el polígono de frecuencias. Responda:

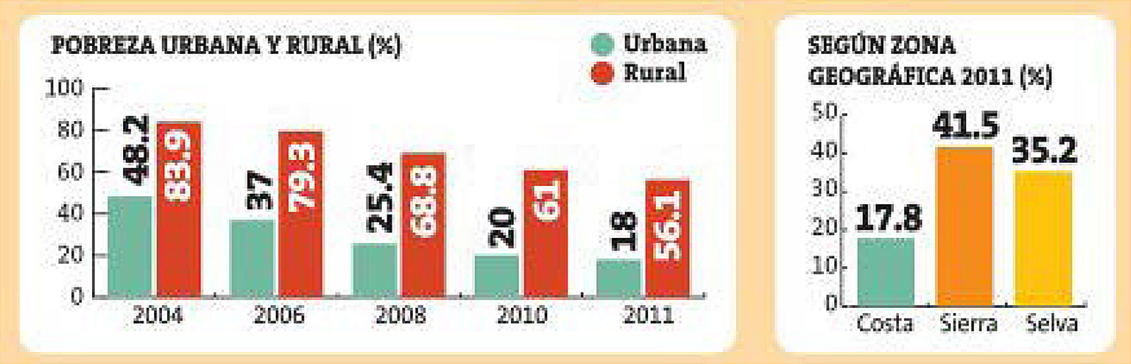
- ¿Se puede afirmar que es una distribución simétrica?
- ¿Hacia qué lado se prolonga más el polígono (izquierda o derecha)?

3. Complete en los espacios en blanco la palabra que otorgue sentido a las siguientes expresiones:

- En el nivel de medición ______________________________ los datos sólo pueden ser clasificados manteniendo un esquema de orden.
- En el nivel de medición de razón, el cero se considera como ____________________________, lo cual significa la ausencia total de la característica en estudio.
- Se llama _________________________________ a aquel elemento que posee la característica en estudio.
- Se llama _________________________________ a aquel número que describe alguna característica de la población.
- La estadística _________________________________ es aquella que partiendo de los datos de la muestra busca sacar conclusiones y tomar decisiones respecto a la población.

4. Según el diario "Perú 21", en su publicación del 31 de mayo del 2012:

"La reducción de la pobreza sigue a un ritmo acelerado. En los últimos cuatro años se redujo en 14 puntos porcentuales al pasar de 42.4% (2007) a 27.8% (2011), según el Instituto Nacional de Estadística e Informática (INEI). De acuerdo con el nuevo cálculo, unos 790 mil peruanos dejaron de ser pobres solo el año pasado. Sin embargo, aún existen 8.3 millones que permanecen en esa situación, precisó Alejandro Vílchez, jefe del INEI, durante la presentación del documento Evolución de la Pobreza 2007-2011. Explicó que la canasta básica per cápita es de S/.272 mensuales, con la cual se establece la línea de pobreza. Si una persona gasta menos de ese monto al mes se le considera pobre. Si el consumo se ubica por debajo de S/.143 se le considera pobre extremo."

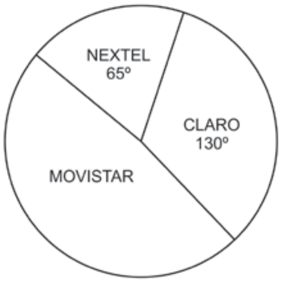


Población peruana en el 2011: 29 797 694 habitantes.

En relación a dicha publicación determine el valor de verdad de las siguientes proposiciones: (4 puntos)

- La población considerada pobre del ámbito rural el año 2011 fue de 16 716 506 habitantes, aproximadamente ( )
- El número de pobladores pobres del ámbito rural disminuyó en 1 460 087 del año 2010 al año 2011 aproximadamente ( )
- En el texto, al mencionar el sueldo de S/.272 mensuales se están refiriendo a un parámetro ( )
- En la sierra, el año 2011 existían 17 431 651 personas que no eran consideradas pobres, aproximadamente ( )

5. El siguiente gráfico de sectores registra la información sobre las preferencias por un operador móvil, de una muestra de 12 000 personas arequipeñas.



- ¿Cuántas personas manifestaron ser clientes de Movistar? ____________________ (2 puntos)
- ¿Qué porcentaje de personas de la muestra prefieren a Nextel? ____________________ (2 puntos)

GLOSARIO DE LA UNIDAD

A

Arreglo u ordenamiento de datos: Organización de los datos sin procesar en orden ascendente o descendente.

C

Conjunto de datos: Una colección de datos.

D

Distribución de frecuencias: Presentación de un conjunto de datos en el que se muestra la frecuencia absoluta y/o relativa de los datos pertenecientes a cada intervalo.

E

Estadígrafo: Característica de la muestra.

F

Frecuencia absoluta: Número de datos que pertenecen a determinado intervalo.

Frecuencia relativa: Proporción representativa de cada intervalo respecto al tamaño total.

H

Histograma: Gráfica de un conjunto de datos compuesta por una serie de rectángulos cada uno con un ancho proporcional al rango de los valores de cada clase o intervalo y altura proporcional a la frecuencia.

I

Intervalo: Conjunto de datos numéricos establecidos entre dos límites, inferior e inferior.

M

Marca de clase: Punto medio de cada intervalo, valor representativo de todos los datos que pertenecen a dicho intervalo.

Muestra: Colección de algunos elementos, subconjunto de la población con las mismas características que la población bajo estudio, utilizada para describir a la población de la cual proviene.

O

Ojiva: Gráfica de una distribución de frecuencias acumuladas.

P

Parámetro: Característica de la población.

Población: Colección de todos los lementos que se están estudiando y sobre los cuales intentamos llegar a conclusiones.

V

Variable cualitativa nominal: Aquella que asume valores cualitativos que no poseen jerarquías entre si.

Variable cualitativa ordinal: Aquella que asume valores cualitativos los cuáles si poseen jerarquías entre si.

Variable cuantitativa continua: Aquella que asume valores cuantitativos que provienen de una medición.

Variable cuantitativa discreta: Aquella que asume valores cuantitativos que provienen de un conteo.

BIBLIOGRAFÍA DE LA UNIDAD I

Devore, J. (2008). *Probabilidad y Estadística (*7 ed.). México: Editorial Cengage Learning Editores S.A.

Martínez Bencardino, C. (2012). *Estadística y muestreo* (13 ed.). Bogotá, Colombia: Editorial Ecoe Ediciones.

Toma, J. (2012). *Estadística aplicada – Primera parte* (2 Ed.). Perú: Fondo editorial de la Universidad del Pacífico

Triola, M.(2009). *Estadística.* México: Pearson Educación.

AUTOEVALUACION N° 1

1. Un conjunto de datos contiene los gastos realizados en publicidad por los 12 candidatos que postularon a la alcaldía de la provincia de Huancayo en las últimas elecciones. De ellos se pudo determinar que el gasto publicitario promedio fue de $ 9 000 por candidato.

En relación a dicho estudio determine la veracidad (V) o falsedad (F) de las siguientes proposiciones:

1.1. El conjunto de mediciones es una población. ( )

1.2. La variable que se mide es "Número de candidatos". ( )

1.3. El valor $ 9 000 es un parámetro. ( )

1.4. La variable en estudio es del tipo cuantitativa continua. ( )

1.5. La muestra estuvo conformada por 12 personas. ( )

1.6. La variable se encuentra en el nivel de medición de razón. ( )

1.7. La unidad estadística es cada uno de los candidatos postulantes. ( )

1.8. El estudio puede ser considerado como un censo. ( )

2. A los clientes de un banco se les pide rellenen un cuestionario en el que se hicieron las siguientes preguntas,

	¿CUÁL ES SU EDAD? (CONSIDERANDO AÑOS, MESES Y DÍAS)
	¿Es usted varón o mujer?
	¿Hace cuánto tiempo empezó a utilizar los servicios del banco? (considerando años, meses y días)
	¿Cuál es su tiempo de servicio? (en años cumplidos)
	¿Cómo califica el servicio que le brindamos? (Opciones: Muy bueno, bueno, regular, deficiente, muy deficiente)

¿Cuáles proporcionan datos cuantitativos o cualitativos? Asigne el código correspondiente:

A= Cualitativos Nominales B= Cualitativos Ordinales

C= Cuantitativos Discretos D=Cuantitativos Continuos

A) DADCB B) DBCAB C) CADCA D) ADBCB E) BCBAC

3. Una marca de cloro líquido se vende en botellas cuya etiqueta dice contener 128 onzas (un galón). Debido a múltiples quejas recibidas de consumidores, INDECOPI decide investigar si la cantidad promedio en las botellas es realmente 128 onzas. En su puesto de inspector de INDECOPI decide visitar algunos comercios y compra 100 botellas de esta marca de cloro para corroborar las quejas de los consumidores. El resultado indica que la cantidad promedio en las botellas es de 126 onzas.

Indique en términos del problema cuál es:

- La población.
- La muestra (indicar su tamaño).
- La variable que se debe medir.
- El parámetro.
- El estadístico.

4. Suponga que una empresa ha llevado a cabo un estudio para analizar el número de microcomputadoras que existen en pequeñas empresas. Para tal fin, toma una muestra de empresas y encuentra los siguientes resultados sobre el número de microcomputadoras que tienen en uso:

5	7	9	8	5	2	4	3	7	8	4	9	6	8	7	6	9	8	4	8
6	4	7	4	3	5	8	5	9	6	7	9	4	7	5	8	7	9	6	8

Luego de organizar los datos en una tabla de frecuencias, es posible afirmar que:

A) Las empresas que tienen de 2 a 4 computadoras representan el 38% de la muestra.

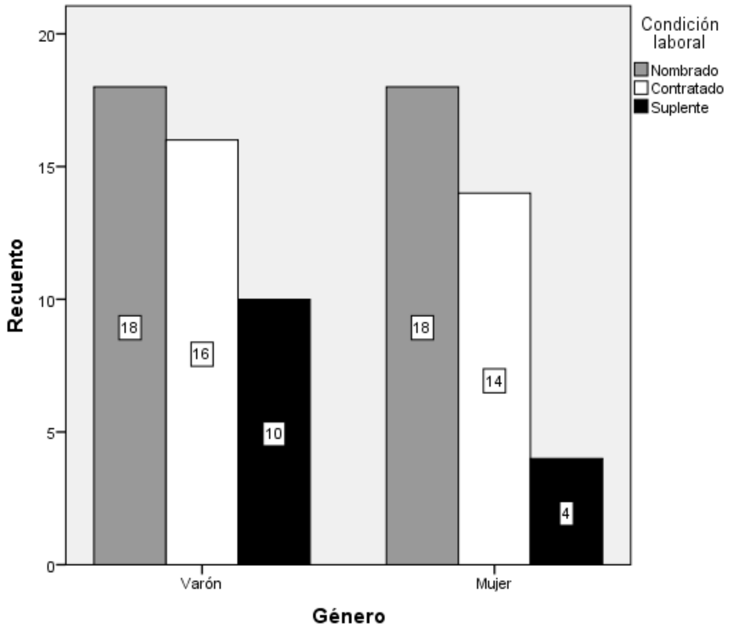
B) La variable en estudio es cuantitativa continua.

C) La muestra estuvo conformada por 38 empresas.

D) La mayoría de las empresas cuentan con 6 computadoras.

E) Las empresas que tienen 9 computadoras representan el 25% de la muestra.

5. Una muestra tomada entre los docentes que participaron en el último concurso para contratas y nombramientos del sector educación, fueron evaluados mediante una prueba de conocimientos conocer el dominio de su materia. Luego de efectuado el proceso los resultados se presentaron en el siguiente gráfico.



Es posible afirmar que:

A. La muestra estuvo conformada por 83 docentes.

B. Las mujeres representan el 34% del total de la muestra.

C. Los docentes varones contratados representan el 20% de la muestra.

D. En la muestra se puede contar en total 16 docentes suplentes.

E. En total hay 34 docentes nombrados.

UNIDAD II

"MEDIDAS ESTADÍSTICAS"

DIAGRAMA DE PRESENTACIÓN DE LA UNIDAD II

RESULTADO DE APRENDIZAJE

Al finalizar la unidad, el estudiante será capaz de interpretar las medidas estadísticas de una encuesta en una organización.

CONTENIDOS	ACTIVIDADES FORMATIVAS (HABILIDADES Y ACTITUDES)	SISTEMA DE EVALUACIÓN (TÉCNICAS Y CRITERIOS)
TEMA N° 1: ESTADÍSTICOS PARA DESCRIBIR, EXPLORAR Y COMPARAR DATOS: 1 Medidas Descriptivas 2 Medidas de tendencia Central 3 Medidas de Dispersión 4 Medidas de posición relativa. 5 Medidas de asimetría y curtosis	· Calcula las medidas de tendencia central y las interpreta. · Calcula las medidas de dispersión y las interpreta. · Utiliza herramientas estadísticas para analizar un conjunto de datos. · Resuelve las actividades de la guía de práctica del texto, haciendo uso del Laboratorio de cómputo.	Procedimientos e indicadores de evaluación permanente · Entrega puntual de trabajos realizados. · Calidad, coherencia y pertinencia de contenidos desarrollados. · Prueba teórico-práctica, individual. · Actividades desarrolladas en sesiones tutorizadas Criterios de evaluación para el análisis estadístico de la encuesta · El estudiante selecciona correctamente el tipo de tabla de frecuencias correspondiente a cada tipo de variable · El estudiante elabora el gráfico adecuado para representar diferentes tipos de variables en diferentes niveles de medición. · El estudiante calcula el estadígrafo correspondiente para resumir diferentes tipos de variables en diferentes niveles de medición. · El estudiante interpreta correctamente los estadígrafos calculados

RECURSOS:

Videos:

Tema N° 1

Media, Varianza y Desviación Estándar usando la calculadora CASIO fx-82MS para datos simples

https://www.youtube.com/watch?v=qguhqq0xvM0

Tema N° 2

Media, Varianza y Desviación Estándar con tablas de frecuencia y calculadora CASIO fx-82MS

https://www.youtube.com/watch?v=2FPsN0oZsyU

Tema N° 3

Media, Varianza y Desviación Estándar de datos simples usando la calculadora CASIO fx-82ES-Plus

https://www.youtube.com/watch?v=7s77eMDRkdc

Tema N° 4

Calcular Media, Varianza y Desviación Estándar con calculadora CASIO fx-350ES

https://www.youtube.com/watch?v=vX0c3M1qZls

Lectura complementaria:

Lectura Seleccionada N° 1

¿Para qué sirve la mediana, si ya tenemos la media aritmética?

Roberto Behar Gutiérrez, Pere Grima Cintas, 2004. España.

Lectura Seleccionada N° 2

¿Cuándo conviene utilizar boxplots para analizar o describir datos?

Roberto Behar Gutiérrez, Pere Grima Cintas, 2004. España

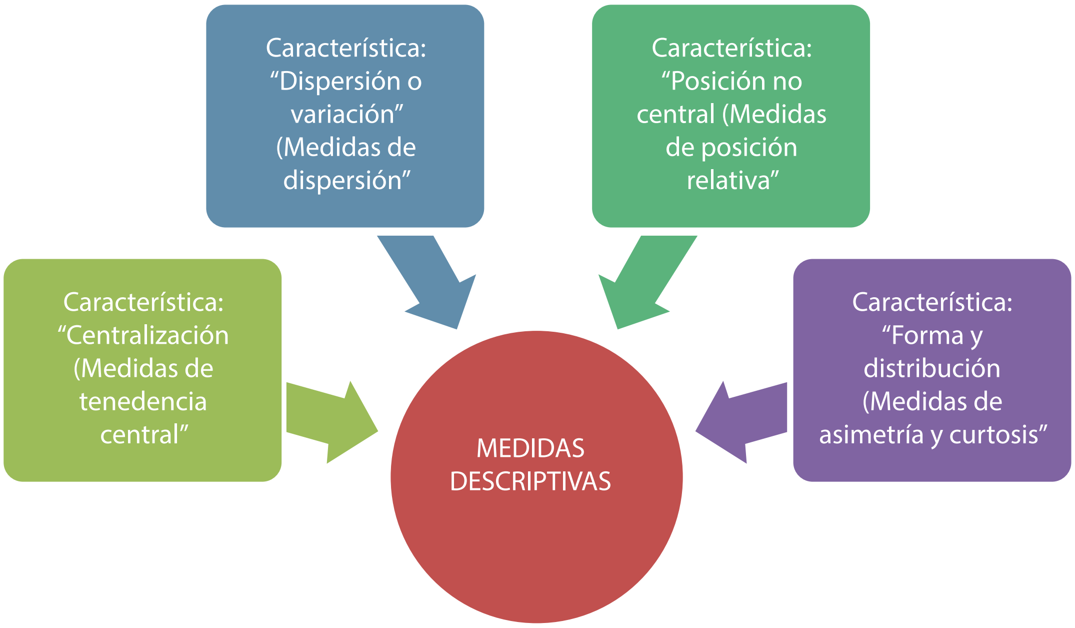
INSTRUMENTO DE EVALUACIÓN	Prueba de desarrollo
BIBLIOGRAFÍA (BÁSICA Y COMPLEMENTARIA)	**BÁSICA** Devore, J. (2008). *Probabilidad y Estadística.* (7 ed.). México: Editorial Cengage Learning Editores S.A. **COMPLEMENTARIA** Ross, S. (2001). *Probabilidad y Estadística para Ingenieros.* (3 ed.). México: Mc Graw Hill. Berenson, M. & Levine, D. (2010). *Estadística Básica en Administración, conceptos y aplicaciones.* México: Prentice Hall.
RECURSOS EDUCATIVOS DIGITALES	· Instituto Nacional de Estadística e Informática. Disponible en: http://www.inei.gob.pe/ · Tokio explora en un congreso la “imprescindible” ciencia de la estadística. (2012, Jul 04). EFE News Service. Disponible en: http://search.proquest.com/docview/1023206430?accountid=146219 · Canales, E. (2005, Jul 26). Mexicar / AMLO sin estadistica. El Norte. Disponible en:: http://search.proquest.com/docview/311786309?accountid=146219

TEMA Nº 1:
ESTADÍSTICOS PARA DESCRIBIR, EXPLORAR Y COMPARAR DATOS.

Estimado estudiante, es momento de ingresar en uno de los temas más importantes y apasionantes de la Estadística Descriptiva. Ahora veremos la manera de calcular medidas que describan a todo un conjunto de datos con un único valor, el cual será considerado representativo. Esta descripción se hará en función a ciertas características que presenten los datos. Aquí es importante el cálculo y la interpretación de dichas medidas.

1. MEDIDAS DESCRIPTIVAS:

Son valores numéricos calculados a partir de la muestra y que nos resumen la información contenida en ella, respecto a ciertas características.



Fuente: Elaboración propia

2. MEDIDAS DE TENDENCIA CENTRAL:

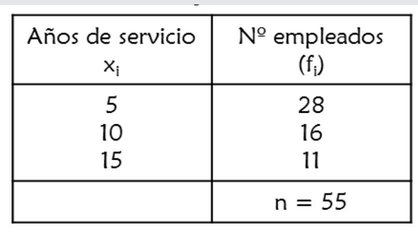
Son valores numéricos que señalan un tipo de “centro” de un conjunto de datos, el cual se emplea para representar a todo el conjunto. La más importante de ellas es la media aritmética, seguida por la mediana.

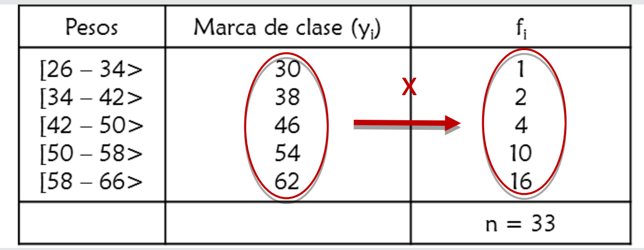
2.1. Media aritmética: $(\mu \ , \ \bar{x})$

Es la medida de tendencia central que se calcula al sumar los valores y dividir el total entre el número de valores. Es común referirse a ella simplemente como la “media”.

Como dice Devore (2008) la media es la medida más conocida y útil. Como casi siempre se pensará que los números x_i constituyen una muestra, a menudo se hará referencia al promedio aritmético como la media muestral.

<table>
<tr><th colspan="3">MEDIA ARITMÉTICA PARA DATOS NO AGRUPADOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$</td>
<td>Las estaturas de una muestra de 5 personas son:
1,65 – 1,58 – 1,60 – 1,72 – 1,70.
Calcule la media de las estaturas.
SOLUCIÓN
$$\overline{x} = \frac{1,65+1,58+1,60+1,72+1,70}{5}$$
$$\overline{x} = 1,65\ m.$$</td>
<td>En promedio las personas tienen una estatura de 1,65 m.</td>
</tr>
</table>

<table>
<tr><th colspan="3">MEDIA ARITMÉTICA PARA DATOS AGRUPADOS DISCRETOS</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>$$\overline{x} = \frac{\sum_{i=1}^{n} x_i . f_i}{n}$$</td>
<td>Calcule la media aritmética de los años de servicio de una muestra de trabajadores.
<table><tr><th>Años de servicio x_i</th><th>Nº empleados (f_i)</th></tr><tr><td>5</td><td>28</td></tr><tr><td>10</td><td>16</td></tr><tr><td>15</td><td>11</td></tr><tr><td></td><td>n = 55</td></tr></table>SOLUCIÓN
$$\overline{x} = \frac{5(28)+10(16)+15(11)}{55}$$
$$\overline{x} = 8,5\, años$$</td>
<td>En promedio, los trabajadores tienen 8,5 años de servicio.</td>
</tr>
</table>

<table>
<tr><th colspan="3">MEDIA ARITMÉTICA PARA DATOS AGRUPADOS CONTINUOS</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>$$\overline{x} = \frac{\sum_{i=1}^{n} y_i . f_i}{n}$$</td>
<td>Calcule la media aritmética de los pesos de una muestra de pacientes de un hospital.
<table><tr><th>Pesos</th><th>Marca de clase (y_i)</th><th>f_i</th></tr><tr><td>[26 – 34></td><td>30</td><td>1</td></tr><tr><td>[34 – 42></td><td>38</td><td>2</td></tr><tr><td>[42 – 50></td><td>46</td><td>4</td></tr><tr><td>[50 – 58></td><td>54</td><td>10</td></tr><tr><td>[58 – 66></td><td>62</td><td>16</td></tr><tr><td></td><td></td><td>n = 33</td></tr></table>SOLUCIÓN
$$\overline{x} = \frac{30(1)+38(2)+46(4)+54(10)+62(16)}{33}$$
$$\overline{x} = 55,2\ kg.$$</td>
<td>El peso promedio de los pacientes es de 55,2 kg.</td>
</tr>
</table>

Propiedades de la media aritmética:

- La suma de las desviaciones (diferencias) entre los valores de la variable y su media aritmética siempre es cero.
- Para un conjunto dado de observaciones, la media aritmética es única.
- La media se encuentra afectada por los valores extremos del conjunto de datos. Si un valor se modifica, la media también se modifica.
- Si a los valores de una variable se le suma o resta una constante, la media aritmética quedará aumentada o disminuida en dicho valor.
- Si a los valores de una variable se le multiplica o divide por una constante, la media aritmética quedará multiplicada o dividida por dicho valor.

2.2. Mediana (Me):

La palabra "mediana" es sinónimo de "medio" y la mediana muestral es en realidad el valor medio (el que ocupa la posición central) una vez que se ordenan las observaciones de la más pequeña a la más grande.

MEDIANA PARA DATOS NO AGRUPADOS (CANTIDAD IMPAR DE DATOS):

Fórmula	Aplicación	Interpretación
$Lugar: \frac{n+1}{2}$	Las estaturas de una muestra de 5 personas (n=5) son: 1,65 – 1,58 – 1,60 – 1,72 – 1,70. Calcule la mediana de sus estaturas. **SOLUCIÓN** PRIMER PASO: Ordenar los datos de menor a mayor. 1,58 – 1,60 – 1,65 – 1,70 – 1,72 SEGUNDO PASO: Identificar el valor que se encuentra en el centro del conjunto de datos. Ese valor es la mediana. **1,58 – 1,60 – 1,65 – 1,70 – 1,72** Me = 1,65	50% de las personas miden a lo más 1,65 m. y el otro 50% miden más de 1,65 m.

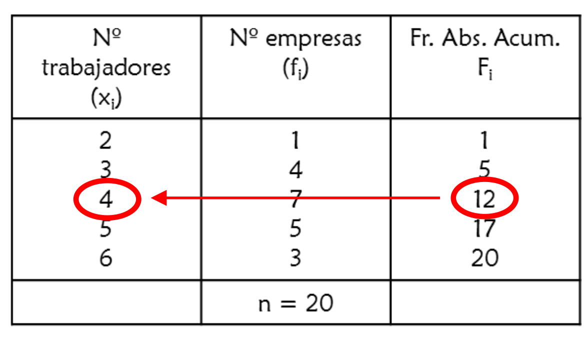
MEDIANA PARA DATOS NO AGRUPADOS (CANTIDAD PAR DE DATOS):

Fórmula	Aplicación	Interpretación
$Lugares$ $\frac{n}{2}\ y\ \frac{n}{2}+1$	Las estaturas de una muestra de 4 personas (n=4) son: 1,65 – 1,60 – 1,72 – 1,70. Calcule mediana de sus estaturas. **SOLUCIÓN** PRIMER PASO: Ordenar los datos de menor a mayor. 1,60 – 1,65 – 1,70 – 1,72 SEGUNDO PASO: Identificar los valores que se encuentran en el centro del conjunto de datos. El promedio de esos valores es la mediana. **1,60 – 1,65 – 1,70 – 1,72** Me = (1,65 + 1,70) / 2 = 1,675 m	50% de las personas miden a lo más 1,675 m. y el otro 50% miden más de 1,675 m.

MEDIANA PARA DATOS AGRUPADOS DISCRETOS: CASO 1

Fórmula | **Aplicación** | **Interpretación**

Calcule la mediana del número de trabajadores de una muestra de 20 empresas.



Nº trabajadores (x_i)	Nº empresas (f_i)	Fr. Abs. Acum. F_i
2	1	1
3	4	5
4	7	12
5	5	17
6	3	20
	n = 20	

SOLUCIÓN

PRIMER PASO: Calcule n/2 (20/2 = 10)

SEGUNDO PASO: Calcule la menor frecuencia absoluta acumulada que supera a n/2 (En este caso 12)

TERCER PASO: Ubicar a la mediana (está dada por el valor xi correspondiente a la frecuencia acumulada que superó a n/2)

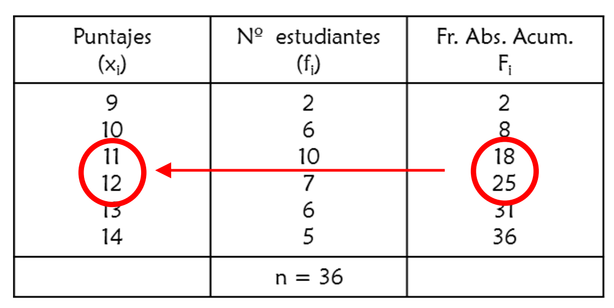
Me = 4

A lo más 50% de las empresas tienen como máximo 4 trabajadores y el otro 50% tiene más de 4 trabajadores

MEDIANA PARA DATOS AGRUPADOS DISCRETOS: CASO 2

Fórmula | **Aplicación** | **Interpretación**

Calcule la mediana de los puntajes de una prueba, de una muestra de 36 estudiantes.



Puntajes (x_i)	Nº estudiantes (f_i)	Fr. Abs. Acum. F_i
9	2	2
10	6	8
11	10	18
12	7	25
13	6	31
14	5	36
	n = 36	

SOLUCIÓN

PRIMER PASO: Calcule n/2 (36/2 = 18)

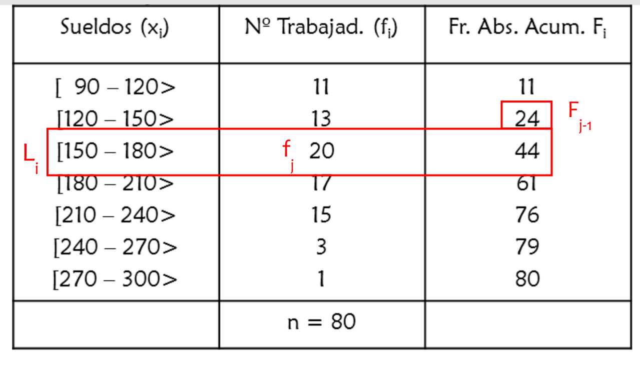
SEGUNDO PASO: Calcule la frecuencia absoluta acumulada que es igual a n/2 y a la siguiente (En este caso 18 y 25)

TERCER PASO: Ubicar a la mediana (está dada por el promedio de los valores xi correspondiente a las frecuencias acumuladas elegidas)

Me = (11 + 12)/2 = 11,5

A lo más el 50% de los estudiantes obtuvo 11,5 puntos y el otro 50% obtuvo más de 11,5 puntos.

<table>
<tr><th colspan="3">MEDIANA PARA DATOS AGRUPADOS CONTINUOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>$$Me = L_i + A\left(\frac{\frac{n}{2}-F_{j-1}}{f_j}\right)$$</td>
<td>
Calcule la mediana de los sueldos de una muestra de 80 trabajadores



<table>
<tr><th>Sueldos (x_i)</th><th>Nº Trabajad. (f_i)</th><th>Fr. Abs. Acum. F_i</th></tr>
<tr><td>[90 – 120></td><td>11</td><td>11</td></tr>
<tr><td>[120 – 150></td><td>13</td><td>24</td></tr>
<tr><td>[150 – 180></td><td>20</td><td>44</td></tr>
<tr><td>[180 – 210></td><td>17</td><td>61</td></tr>
<tr><td>[210 – 240></td><td>15</td><td>76</td></tr>
<tr><td>[240 – 270></td><td>3</td><td>79</td></tr>
<tr><td>[270 – 300></td><td>1</td><td>80</td></tr>
<tr><td></td><td>n = 80</td><td></td></tr>
</table>

SOLUCIÓN

PRIMER PASO: Calcule n/2 (80 / 2 = 40)

SEGUNDO PASO: Ubique el intervalo mediano (el que tenga frecuencia acumulada inmediatamente superior a n/2. En este caso 44)

Li = 150
A = 180 – 150 = 30
n/2 = 40
Fj-1 = 24
fj = 20

TERCER PASO: Calcule la mediana aplicando la fórmula.

$$Me = 150 + 30\left(\frac{40-24}{20}\right) = 174$$
</td>
<td>A lo más el 50% de los sueldos son como máximo de S/. 174 y el otro 50% de sueldos son mayores de S/. 174</td>
</tr>
</table>

Propiedades de la Mediana

- La mediana no se encuentra afectada de valores extremos, sino por la cantidad de observaciones.
- Se emplea como reemplazo de la media aritmética cuando la distribución presenta valores muy alejados.
- Es útil cuando se tienen datos cualitativos susceptibles de ordenarse de acuerdo a rangos, calificaciones o categorías (ejemplo: Muy malo, malo, regular, bueno, muy bueno)
- La mediana no es adecuada para manipulaciones algebraicas.

2.3. Moda (Mo):

Como menciona Martínez Bencardino (2012), es el valor de la variable que más se repite o aquel que presenta la máxima frecuencia. Es la única medida que puede ser empleada en datos cualitativos. Un conjunto de datos puede ser “amodal” (sin moda), “unimodal” (una sola moda), “bimodal” (dos modas) o “multimodal” (más de dos modas)

<table>
<tr><th colspan="3">MODA PARA DATOS NO AGRUPADOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td></td>
<td>
Las estaturas de una muestra de 9 personas son: 1,60 – 1,75 – 1,70 – 1,65 – 1,58 – 1,60 – 1,60 – 1,60 – 1,58 – 1,70. Calcule la moda de sus estaturas.

SOLUCIÓN

1,60 – 1,75 – 1,70 – 1,65 – 1,58 – 1,60 – 1,60 – 1,60 – 1,58 – 1,70

Por simple inspección:

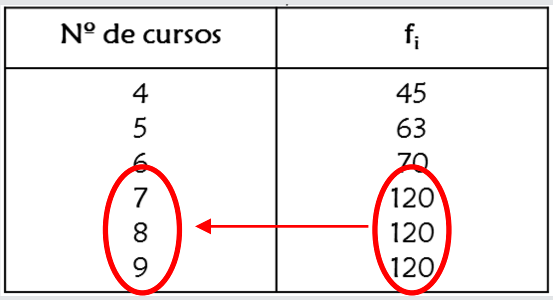
Mo = 1,60 m.
</td>
<td>La estatura que más se repite es 1,60 m. (Unimodal)</td>
</tr>
</table>

MODA PARA DATOS AGRUPADOS: VARIABLE CUANTITATIVA DISCRETA

Fórmula

Aplicación

Se realizó una encuesta a un grupo de estudiantes de la Universidad Continental, para averiguar en cuántos cursos se han matriculado para el semestre académico 2015 – II. Calcule la moda del número de cursos en que se matricularon.



Nº de cursos	f_i
4	45
5	63
6	70
7	120
8	120
9	120

SOLUCIÓN

PRIMER PASO: Identificar la frecuencia absoluta más alta.

SEGUNDO PASO: Elija la moda (observación que más se repite).

Mo1 = 1, Mo2 = 8, Mo3 = 9

Interpretación

La cantidad de cursos matriculados más frecuente entre los estudiantes es de 7; 8 y 9 cursos.

MODA PARA DATOS AGRUPADOS: VARIABLE CUANTITATIVA CONTINUA

Fórmula

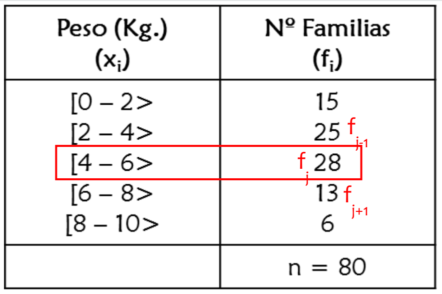
$$Mo = L_i + A\left(\frac{d_1}{d_1 + d_2}\right)$$

$$d_1 = f_j - f_{j-1}$$

$$d_2 = f_j - f_{j+1}$$

Aplicación

En una encuesta en un hospital, se indagó acerca de la cantidad de carne roja que consumen un grupo de 80 pacientes (en kg). Calcule la moda del consumo de carne.



Peso (Kg.) (x_i)	Nº Familias (f_i)
[0 – 2>	15
[2 – 4>	25 f_{j-1}
[4 – 6>	f_j 28
[6 – 8>	13 f_{j+1}
[8 – 10>	6
	n = 80

SOLUCIÓN

PRIMER PASO: Identificar la frecuencia absoluta más alta (28)

SEGUNDO PASO: Identificar el intervalo modal (en este caso el que corresponde a la frecuencia 28) y calcule la moda con la fórmula.

$$\left.\begin{array}{l} f_{j-1} = 25 \\ f_{j+1} = 13 \end{array}\right\} \begin{array}{l} d_1 = 28 - 25 = 3 \\ d_2 = 28 - 13 = 15 \end{array}$$

$$Li = 4$$

$$A = 6 - 4 = 2$$

$$Mo = 4 + 2\left(\frac{3}{3 + 15}\right) = 4{,}3$$

Interpretación

La cantidad de carne que mayormente consumen los pacientes es de 4,3 kg.

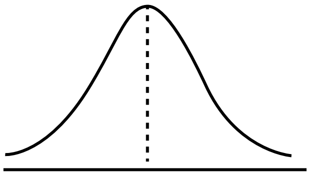
Propiedades de la Moda:

- Se puede utilizar como una localización tanto para datos cualitativos como cuantitativos.

- No está afectada por los valores extremos.
- Muy a menudo no hay un solo valor modal. Pueden darse distribuciones unimodales (1 moda), bimodales (2 modas), multimodales (más de 2 modas), las cuales son difíciles de interpretar.
- Es una medida inestable porque varía si se cambia el intervalo de clase (es decir, si varía la amplitud)
- No se presta a manipulaciones algebraicas.

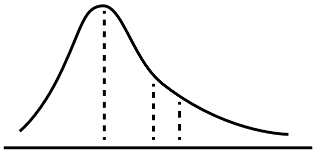
2.4. Relación entre las medidas de tendencia central:

Si la distribución de frecuencias de los datos es simétrica, entonces la media, la mediana y la moda tienen el mismo valor:



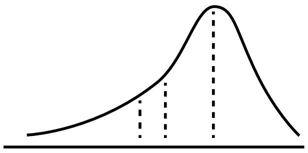
$\bar{X}$ = Me = Mo

Si la distribución es asimétrica de cola a la derecha, entonces, la moda es menor que la mediana y ésta a su vez es menor que la media:



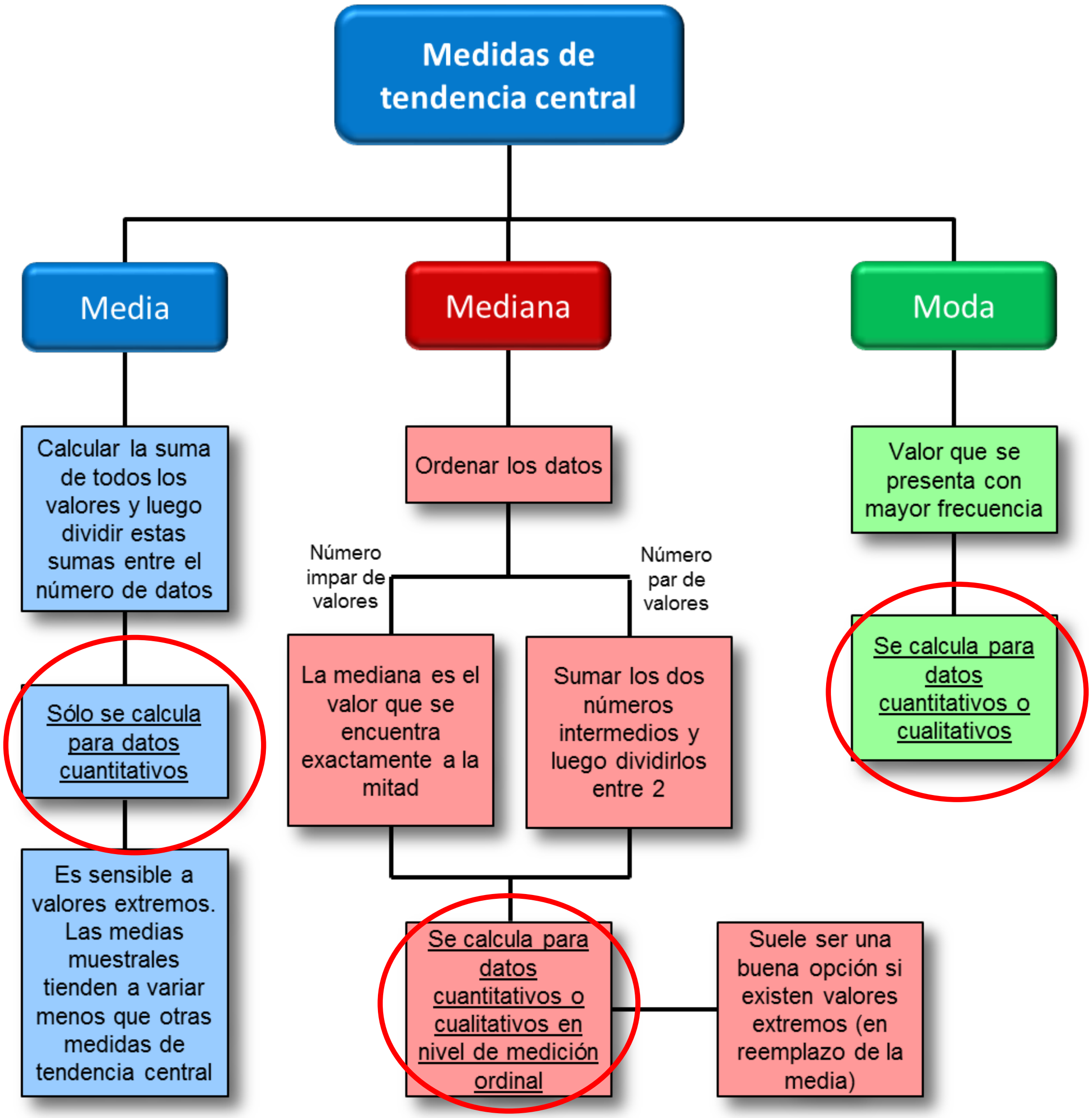
Mo < Me < $\bar{X}$

Si la distribución es asimétrica de cola a la izquierda, entonces la media es menor que la mediana y ésta a su vez es menor que la moda.



$\bar{X}$ < Me < Mo

EN SÍNTESIS



Fuente: Elaboración propia

3. MEDIDAS DE DISPERSIÓN:

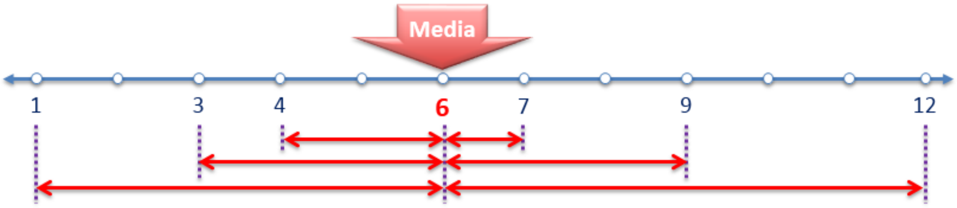
Como menciona Devore (2008) el reporte de una medida de centro da sólo información parcial sobre un conjunto o distribución de datos. Diferentes muestras o poblaciones pueden tener medidas idénticas de centro y aún diferir entre sí en otras importantes maneras.

3.1. Dispersión:

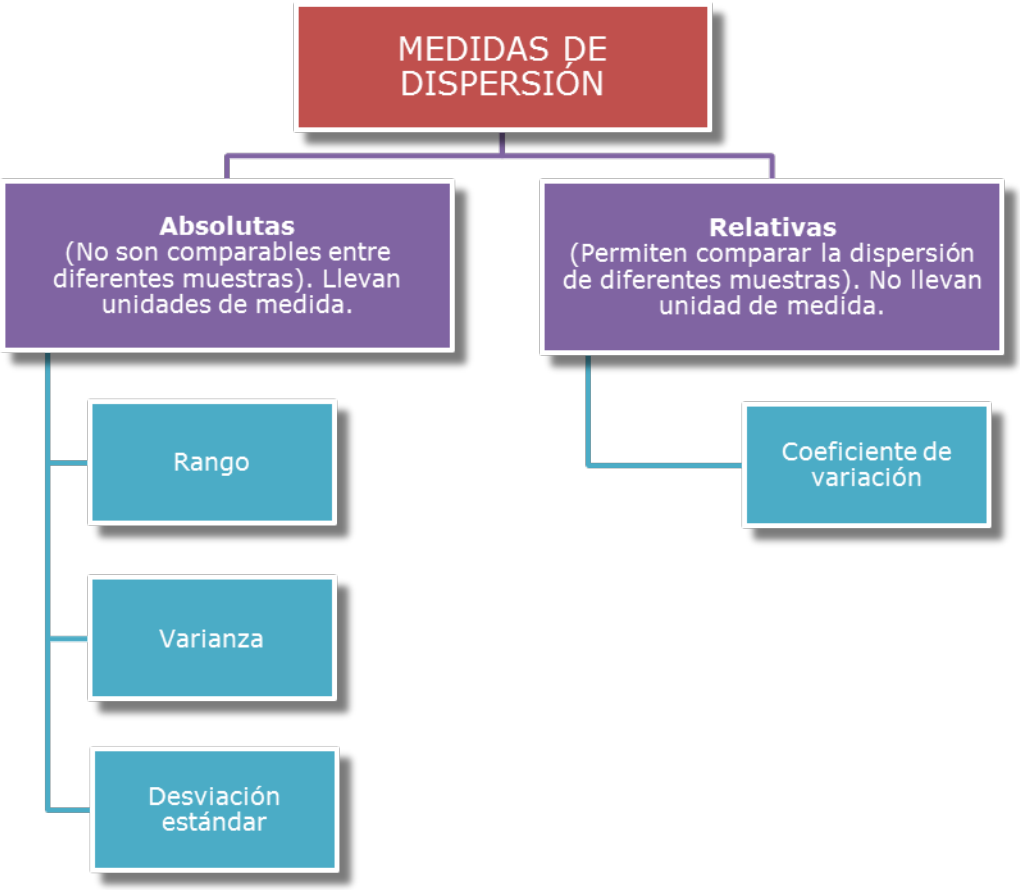
Las medidas de dispersión son valores que sirven para cuantificar la homogeneidad (uniformidad) o dispersión (variabilidad) de los datos, es decir, para medir la proximidad que tienen los datos entre sí.

Datos: 1; 3; 4; 7; 9; 12

Media = 6



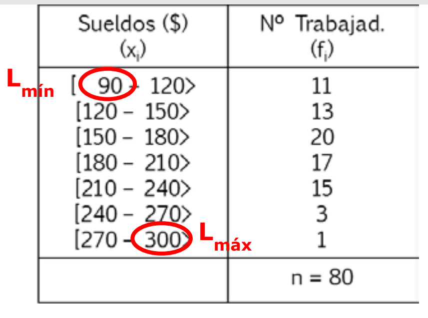
Una medida de dispersión pequeña indica que los datos se acumulan alrededor de la media aritmética. Por consiguiente, la media se considera representativa de los datos. Por el contrario, una medida de dispersión grande indica que la media no es confiable. En tal caso, se asume a la mediana como medida representativa del conjunto de datos.



Fuente: Elaboración propia

3.2. Rango o recorrido (R):

Para Devore (2008) es la medida de dispersión más simple. Representa la diferencia (distancia) entre los valores muestrales más grande y más pequeño.

<table>
<tr><th colspan="3">RANGO PARA DATOS NO AGRUPADOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>R = Máx–Mín</td>
<td>El último fin de semana hubo cinco representantes de servicio al cliente que trabajaron en tiendas CARSA. Las cantidades de televisores que vendieron estos representantes son:
5; 8; 4; 10 y 3
Calcule el rango de las ventas.
SOLUCIÓN
PRIMER PASO: Identifique el valor máximo y el mínimo
Máx = 10 Mín = 3
SEGUNDO PASO: Calcule la diferencia entre el máximo y el mínimo
R = 10 – 3 = 7</td>
<td>La distancia (o diferencia) que existe entre la mayor y menor cantidad de televisores vendidos es de 7 televisores.</td>
</tr>
<tr><th colspan="3">RANGO PARA DATOS AGRUPADOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>R = Lmín – Lmáx</td>
<td>Calcule el rango de los sueldos semanales de una muestra de trabajadores de la Municipalidad Provincial de Huancayo.
<table><tr><th>Sueldos ($)
(x_i)</th><th>N° Trabajad.
(f_i)</th></tr><tr><td>[90 – 120></td><td>11</td></tr><tr><td>[120 – 150></td><td>13</td></tr><tr><td>[150 – 180></td><td>20</td></tr><tr><td>[180 – 210></td><td>17</td></tr><tr><td>[210 – 240></td><td>15</td></tr><tr><td>[240 – 270></td><td>3</td></tr><tr><td>[270 – 300></td><td>1</td></tr><tr><td></td><td>n = 80</td></tr></table>
SOLUCIÓN:
PRIMER PASO: Identifique el valor máximo y el mínimo
Lmáx = 300 Lmín = 90
SEGUNDO PASO: Calcule la diferencia entre el máximo y el mínimo
R = 300 – 90 = 210</td>
<td>La distancia (o diferencia) que existe entre el mayor y menor sueldo semanal es de S/. 210.</td>
</tr>
</table>

Propiedades del Rango:

- Fácil de calcular.
- Siempre asume valores positivos.
- Es la primera medida de dispersión que debe usarse porque permite conocer el intervalo de variación de los datos.
- Una desventaja es que no describe la variabilidad de los datos que se encuentran comprendidos entre los valores mínimo y máximo.

3.3. Varianza (s2 , s2):

Se define como la media aritmética de las desviaciones de los datos respecto a la media, elevadas al cuadrado. Es una de las medidas descriptivas más importantes.

Se puede calcular para poblaciones como para muestras, aunque el cálculo en este último caso difiere un poco del poblacional.

Varianza poblacional

$$s^2=\frac{\sum\left(x-\overline{x}\right)^2}{n-1}$$

Varianza muestral

$$\sigma^2=\frac{\sum\left(x-\mu\right)^2}{N}$$

La diferencia radica en el denominador. Si se utiliza un divisor "n" en la varianza muestral, entonces la cantidad resultante tendería a subestimar a s^2 (se producen valores demasiado pequeños en promedio), mientras que si se divide entre el divisor un poco más pequeño "n – 1" se corrige esta subestimación.

En el cálculo de la varianza las unidades de medida de los datos quedan elevadas al cuadrado, motivo por el cual no se puede interpretar.

3.4. Desviación estándar o típica (s , s)

Es la raíz cuadrada positiva de la varianza.

Desv. Estand. Poblacional

$$s=\sqrt{\frac{\sum\left(x-\overline{x}\right)^2}{n-1}}$$

Desv. Estand. Muestral

$$\sigma=\sqrt{\frac{\sum\left(x-\mu\right)^2}{N}}$$

VARIANZA Y DESVIACIÓN ESTÁNDAR PARA DATOS NO AGRUPADOS:		
Fórmula	**Aplicación**	**Interpretación**
Varianza $s^2=\frac{\sum\left(x-\overline{x}\right)^2}{n-1}$ Desv. Estándar $s=\sqrt{s^2}$	Considere las siguientes edades de 5 niños como una muestra: 8; 3; 7; 3; 4. Calcule la varianza y desviación estándar de dichas edades. SOLUCIÓN Para calcular la varianza: PRIMER PASO: Calcule la media de la muestra $\overline{x}$ = (8+3+7+3+4)/5 = 5 SEGUNDO PASO: Calcule la sumatoria de las desviaciones cuadráticas respecto a la media. (8–5)2 + (3–5)2 + (7–5)2 + (3–5)2 + (4–5)2 =22 TERCER PASO: Divida la sumatoria anterior entre el número de elementos de la muestra disminuido en 1 (n=5) s2 = 22/(5-1) = 5,5 años Para calcular la desviación estándar: Extraiga la raíz cuadrada de la varianza. $s=\sqrt{5{,}5\,\text{años}^2}=2{,}34520788$	La varianza no se puede interpretar por tener sus unidades de medida elevadas al cuadrado. La desviación estándar indica que las edades se dispersan o varían en promedio en 2,3 años respecto a la edad promedio.

<table>
<tr><th colspan="3">VARIANZA Y DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS:</th></tr>
<tr><th>Fórmula</th><th>Aplicación</th><th>Interpretación</th></tr>
<tr>
<td>Varianza

$$s^2 = \frac{\sum (x - \overline{x})^2 . f_i}{n - 1}$$

Desv. Estándar

$$s = \sqrt{s^2}$$</td>
<td>Los sueldos semanales de una muestra de trabajadores de la Municipalidad Provincial de Huancayo se presenta a continuación. Calcule la varianza y la desviación estándar de dichos datos.

<table>
<tr><th>Sueldos</th><th>Marca clase</th><th>Nº. Trabajadores</th></tr>
<tr><td>[90 - 120></td><td>105</td><td>11</td></tr>
<tr><td>[120 - 150></td><td>135</td><td>13</td></tr>
<tr><td>[150 - 180></td><td>165</td><td>20</td></tr>
<tr><td>[180 - 210></td><td>195</td><td>15</td></tr>
<tr><td>[210 - 240></td><td>225</td><td>9</td></tr>
<tr><td></td><td></td><td>n = 68</td></tr>
</table>

SOLUCIÓN

Para calcular la varianza:

PRIMER PASO: Calcule la media de la muestra

$$\overline{x} = \frac{11(105) + \ldots + 9(225)}{68} = 164{,}1176$$

SEGUNDO PASO: Calcule la sumatoria de las desviaciones cuadráticas respecto a la media.

$(105-164{,}1176)^2 x11 + \ldots +$

$(225-164{,}1176)^2 x9$

$= 97147{,}05882$

TERCER PASO: Divida la sumatoria anterior entre el número de elementos de la muestra disminuido en 1 (n=68)

$s^2 = 97147{,}05882 /(68-1) =$ 1449,956102 soles.

Para calcular la desviación estándar:

Extraiga la raíz cuadrada de la varianza.

$$s = \sqrt{1449{,}956102 \text{ Soles}^2} = 38{,}07828911 \text{ soles}$$</td>
<td>La varianza no se puede interpretar por tener sus unidades de medida elevadas al cuadrado.

Los sueldos de los trabajadores tienden a dispersarse respecto a la media en 38,1 Soles en promedio.</td>
</tr>
</table>

Propiedades de la varianza:

- Siempre es positiva.
- Es única y siempre existe.
- Está afectada por los valores extremos de los datos.
- Su unidad de medida es el cuadrado de la unidad de medida de los datos originales.
- La varianza será siempre un valor positivo o cero, en el caso de que las puntuaciones sean iguales.
- Si a todos los valores de la variable se les suma un mismo número la varianza no varía.
- Si todos los valores de la variable se multiplican por un mismo número la varianza queda multiplicada por el cuadrado de dicho número.

Propiedades de la desviación estándar:

- Siempre es positiva.
- Está afectada por los valores extremos de los datos.
- Su unidad de medida es igual a la unidad de medida de los datos originales.
- En el caso de que se desee comparar la variabilidad de dos o más conjuntos, la desviación estándar muestral puede usarse únicamente si los conjuntos por compararse tienen las mismas unidades de medida y si las medias muestrales de los conjuntos de datos tienen valores próximos entre sí.
- Si a todos los valores de la variable se les suma un mismo número la desviación estándar no varía.
- Si todos los valores de la variable se multiplican por un mismo número la desviación estándar queda multiplicada por dicho número.

3.5. Coeficiente de variación (C.V.):

Con frecuencia nos interesa establecer comparaciones de la dispersión, entre diferentes muestras que posean distintas unidades de medida (por ejemplo, Nuevos Soles con Dólares), las medidas de dispersión antes estudiadas no permiten realizar este tipo de comparaciones, pero es el coeficiente de variación quien nos ayuda a realizar estas comparaciones.

El coeficiente de variación es la relación entre la desviación estándar o típica de un conjunto de datos y su media aritmética. Se suele expresar en porcentajes y carece de unidad de medida, permite comparar dos o más distribuciones cuando las unidades de medida de las variables están expresadas en diferentes unidades o escalas de medida (por ejemplo, comparar sueldos expresados en Dólares y Nuevos Soles, superficies expresadas en metros cuadrados y pies cuadrados, estaturas expresadas en metros y pesos en kilogramos, etc.). Se calcula de la misma manera tanto para datos agrupados y no agrupados.

$$CV = \frac{s}{\overline{x}} \times 100$$

Comparando dos o más distribuciones, es más homogénea o presenta menos dispersión, aquella distribución que tiene el menor coeficiente de variación. Y por ello también se puede afirmar que tiene una media más representativa.

Para su interpretación se hace uso de la siguiente tabla:

COEF. VARIAC.	GRADO DE VARIABILIDAD
$0 \leq CV < 10$	Datos muy homogéneos
$10 \leq CV < 15$	Datos regularmente homogéneos
$15 \leq CV < 20$	Datos regularmente variables
$20 \leq CV < 25$	Datos variables
$CV \geq 25$	Datos muy variables

Fuente: Jorge Toma Inafuko

COEFICIENTE DE VARIACIÓN		
Fórmula	**Aplicación**	**Interpretación**
$CV = \frac{s}{\bar{x}} \times 100$	El gerente de la empresa de giros y encomiendas "Cargo 1" está pensando si es conveniente adquirir una nueva flota de camiones. Al guardar los paquetes en los camiones para su entrega se debe tomar en cuenta 2 características principales: El peso (en kilogramos) y el volumen (en pies cúbicos) de cada paquete. En una muestra de 20 paquetes se encuentra que la media del peso es de 26 kilogramos, con una desviación estándar de 3,9 kilogramos, mientras que la media del volumen es de 8,8 pies cúbicos, con una desviación estándar de 2,2 pies cúbicos. ¿Cómo puede el gerente comparar la variación del peso y el volumen? **SOLUCIÓN** PRIMER PASO: Calcule el coeficiente de variación de los pesos. $CV = \frac{3{,}9\text{kg}}{26\text{kg}} \times 100 = 15\%$ SEGUNDO PASO: Calcule el coeficiente de variación de los volúmenes. $CV = \frac{2{,}2\ \cancel{\text{pies cúbicos}}}{8{,}8\ \cancel{\text{pies cúbicos}}} \times 100 = 25\%$ TERCER PASO: Compare ambos coeficientes de variación $CV_{PESO} < CV_{VOLUMEN}$	El peso de las cajas es menos disperso que su volumen. Además, se puede considerar que los pesos son datos regularmente variables, en tanto que los volúmenes son datos muy variables.

3.6. Aplicaciones de la desviación estándar:

Como menciona Triola (2009) entre las aplicaciones de la desviación estándar se puede encontrar:

- Regla empírica para datos con distribución normal:

En cualquier distribución de frecuencias simétrica con forma de campana se cumple que:

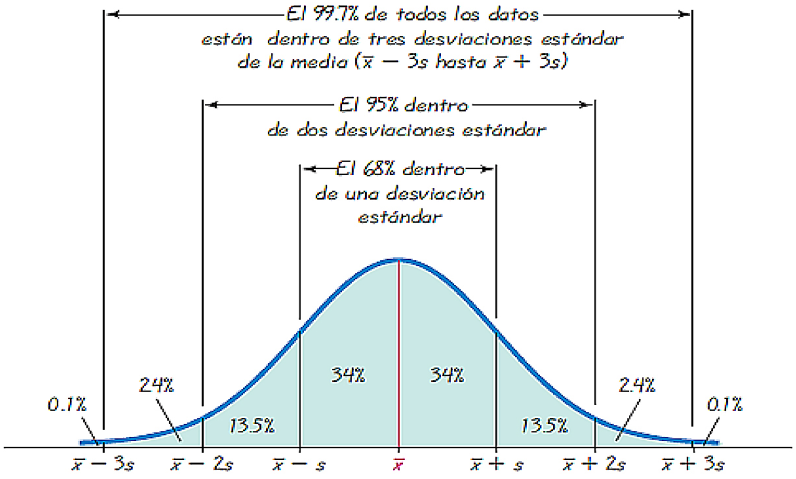


Figura 12. Regla empírica para datos con distribución normal
Fuente: Mario Triola

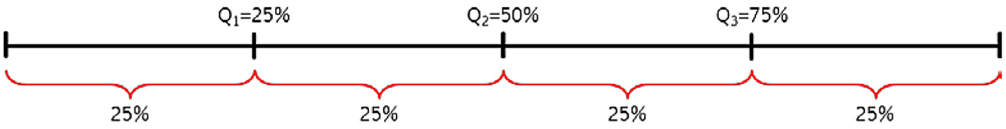
4. MEDIDAS DE POSICIÓN RELATIVA:

Son estadígrafos que dividen a los datos previamente ordenados, en otras proporciones y no sólo en mitades como lo hace la mediana.

Estas medidas se llaman "cuantiles". Los cuantiles más usados en el análisis estadístico son los cuartiles y los percentiles.

4.1. Los cuartiles

Son valores que dividen a un conjunto de observaciones en cuatro partes iguales. Hay tres cuartiles:



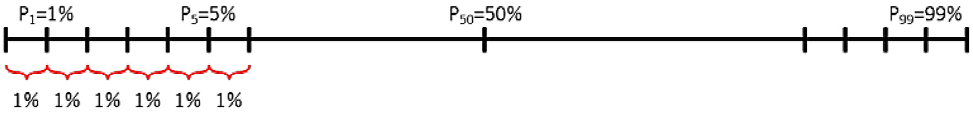
El cuartil 1 (Q_1) es el valor por debajo del cual se encuentra el 25% de las observaciones.

El cuartil 2 (Q_2) es el valor por debajo del cual se encuentra el 50% de las observaciones. Numéricamente es igual a la mediana.

El cuartil 3 (Q_3) es el valor por debajo del cual se encuentra el 75% de las observaciones.

4.2. Los percentiles:

Dividen al conjunto de observaciones en cien partes iguales. Hay 99 percentiles.



El percentil 1 (P_1) es el valor por debajo del cual se encuentra el 1% de las observaciones.

El percentil 25 (P_{25}) es el valor por debajo del cual se encuentra el 25% de las observaciones. Numéricamente es igual que el Q_1.

El percentil 50 (P_{50}) es el valor por debajo del cual se encuentra el 50% de las observaciones. Numéricamente es igual que la mediana y el Q_2.

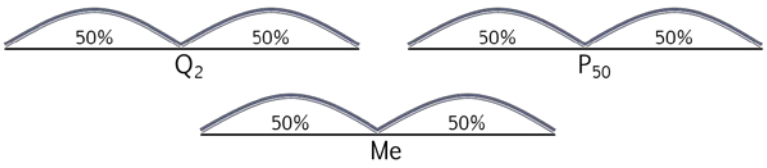
El percentil 75 (P_{75}) es el valor por debajo del cual se encuentra el 75% de las observaciones. Numéricamente es igual que el Q_3.

El percentil 83 (P_{83}) es el valor por debajo del cual se encuentra el 83% de las observaciones.

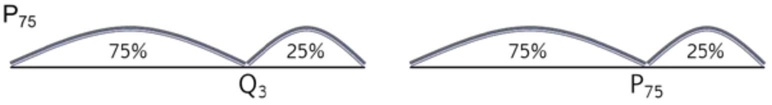
Equivalencias:

$Q_1 = P_{25}$

$Q_2 = P_{50} =$ Mediana



$Q_3 = P_{75}$



PERCENTILES PARA DATOS NO AGRUPADOS: CASO I		
Fórmula	**Aplicación**	**Interpretación**
Lugar : $L_{P_k} = \frac{k}{100} \times n$	Determine el cuartil 1 de las siguientes edades de una muestra de personas: 47; 49; 54; 59; 46: 49; 55; 51; 54; 55; 53 SOLUCIÓN PRIMER PASO: Por equivalencia: Q1 = P25 SEGUNDO PASO: Ordene de menor a mayor los datos (n = 11) 46–47–49–49–51–53–54–54–55–55–59 TERCER PASO: Calcule el lugar para: P25 → k = 25 L = (25/100).11 = 2,75 = 3 (Si el valor de L no es entero, entonces se redondea al entero inmediato superior. La mediana será el dato que ocupa esa posición) CUARTO PASO: Calcule el P25 (Tercer lugar) 46–47–49–49–51–53–54–54–55–55–59 **P25 = 49**	El 25% de los pacientes tienen como máximo 49 años y el 75% restante son mayores de 49 años.

PERCENTILES PARA DATOS NO AGRUPADOS: CASO II

Fórmula

$$Lugar:$$

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

Aplicación

Determine el cuartil 3 de las siguientes edades de una muestra de personas:

25 – 18 – 47 – 35 – 32 – 19 – 20 – 26 – 35 – 30 – 28 – 30

SOLUCIÓN

PRIMER PASO: Por equivalencia: Q3 = P75

SEGUNDO PASO: Ordene de menor a mayor los datos (n = 12)

18–19–20–25–26–28–30–30–32–35–35–47

TERCER PASO: Calcule el lugar para:

P_{75} → k = 75

L = (75/100).12 = 9

(Si el valor de L es entero, entonces se cuenta hasta esa posición y se toma a ese dato y al que le sigue. La mediana será el promedio ambos datos)

CUARTO PASO: Calcule el P75.

18–19–20–25–26–28–30–30–32–35–35–47

P75 = (32 + 35)/2 = 33,5

Interpretación

El 75% de las personas tie-nen como má-ximo 33,5 años y el 25% restante son mayo-res de 33,5 años.

PERCENTILES PARA DATOS AGRUPADOS:

Fórmula

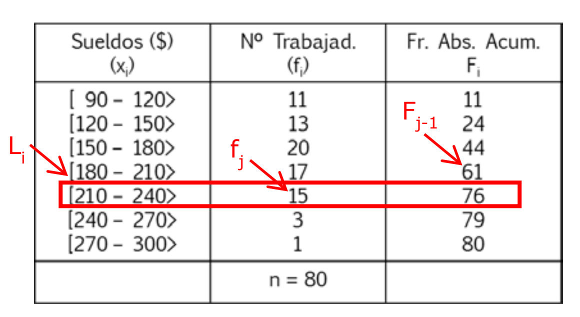
$$Lugar:$$

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

$$P_k = Li + A\left(\frac{\frac{k}{100} \times n - F_{j-1}}{f_j}\right)$$

Aplicación

Dada la siguiente tabla de distribución referida a una muestra de trabajadores, según sueldo semanal; calcule P83 e interprete.



Sueldos ($) ($x_i$)	Nº Trabajad. (f_i)	Fr. Abs. Acum. F_i
[90 – 120>	11	11
[120 – 150>	13	24
[150 – 180>	20	44
[180 – 210>	17	61
[210 – 240>	15	76
[240 – 270>	3	79
[270 – 300>	1	80
	n = 80	

SOLUCIÓN

PRIMER PASO: Calcule el lugar para:

P_{83} → k = 83

L = (83/100).80 = 66,4

SEGUNDO PASO: Busque el valor de L en las frecuencias absolutas acumuladas. De no aparecer, se toma al entero inmediato superior (en este caso 76) y señale el intervalo del P83.

L_i = 210

f_j =15

F_{j-1}=61

A = 240 – 210 = 30

TERCER PASO: Calcule el P83.

$$P_{83} = 210 + 30\left(\frac{66,4 - 61}{15}\right) = 220,8$$

Interpretación

El 83% de los trabajadores tienen como máximo un sueldo de 220,8 Dólares, mientras que el 17% restante tienen sueldos mayores.

4.3. Rango intercuartil:

Es una medida de dispersión estadística, la cual indica la distancia a la que se encuentra el 50% central de datos. Mediante esta medida se eliminan los valores extremadamente alejados. El rango intercuartílico es altamente recomendable cuando la medida de tendencia central utilizada es la mediana (ya que la mediana es insensible a posibles valores extremos)

$$R.I. = Q_3 - Q_1$$

4.4. Análisis exploratorio de datos:

Se utiliza para examinar los conjuntos de datos previamente al trabajo estadístico. De esta manera se consigue obtener un entendimiento del tipo de datos con que se cuenta.

Proporciona métodos sencillos para detectar fallos o deficiencias en el recojo de datos, identificación de datos alejados y comprobación de supuestos.

Valores extremos o datos distantes: Se llama "valor extremo" o "dato distante" a aquel que está muy alejado de la mayor parte de los demás valores. Los valores extremos se deben considerar ya que pueden revelar información importante y afectar en gran medida el valor de la media y de la desviación estándar.

Ejemplo: Edades: 12 – 10 – 13 – 15 – 17 – (46) ➔ Valor extremo

Gráfico de caja y bigote:

Una caja es un rectángulo que se construye sobre la base de los valores del primer cuartil, la mediana y el tercer cuartil. Permite comparar diversos conjuntos de datos simultáneamente respecto a simetría, variabilidad, centro, valores extremos y valores atípicos.

Pasos para su elaboración**:**

Construya un diagrama de caja y bigote con los siguientes datos:

Me =32, Q1 = 28 , Q3 = 38

a) Se dibuja una caja cuyos extremos se localicen en el primer y tercer cuartiles. Esta caja contiene el 50% de los datos centrales.

b) En el punto donde se localiza la mediana se traza una línea.

c) Usando el rango intercuartílico (RI = $Q_3 - Q_1$) se localizan los límites. En un diagrama de caja los límites se encuentran 1,5(RI) abajo del Q_1 y 1,5(RI) arriba del Q_3. Los datos que quedan fuera de estos límites se consideran observaciones atípicas o extremas.

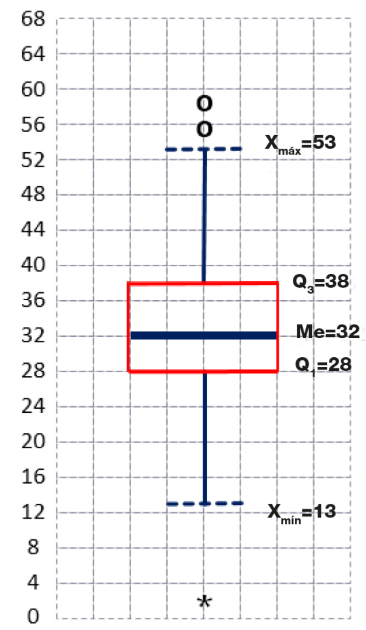
$$RI = 38 - 28 = 10$$

$$Mín = 28 - 1{,}5(10) = 13$$

$$Máx = 38 + 1{,}5(10) = 53$$

d) Los bigotes van desde los extremos de la caja hasta los valores calculados del máximo y mínimo.

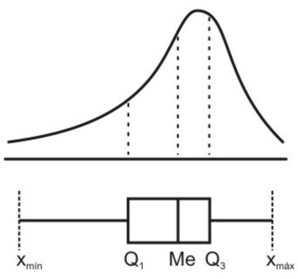
e) Por último, mediante un círculo se indica la localización de las observaciones atípicas y con un asterisco los valores extremos.



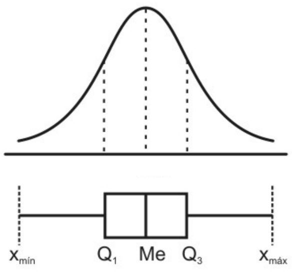
Fuente: Elaboración propia

Levine et al. (2006) presentan la siguiente forma de interpretación del diagrama de caja y bigote:

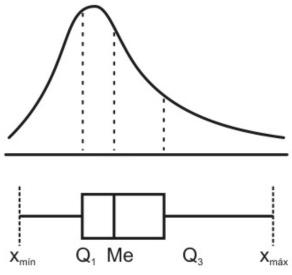
COMPARACIÓN	**TIPO DE DISTRIBUCIÓN**		
	Asimétrico a la izquierda o asimetría negativa	**Simétrico**	**Asimétrico a la derecha o asimetría positiva**
La distancia de $x_{mín}$ a la mediana contra la distancia de la mediana a $x_{máx}$	La distancia de $x_{mín}$ a la mediana es mayor que la distancia de la mediana a $x_{máx}$	Ambas distancias son iguales	La distancia de $x_{mín}$ a la mediana es menor que la distancia de la mediana a $x_{máx}$
La distancia de $x_{mín}$ a Q_1 contra la distancia de Q_3 a $x_{máx}$	La distancia de $x_{mín}$ a Q_1 es mayor que la distancia de Q_3 a $x_{máx}$	Ambas distancias son iguales	La distancia de $x_{mín}$ a Q_1 es menor que la distancia de Q_3 a $x_{máx}$
La distancia de Q_1 a la mediana contra la distancia de la mediana a Q_3	La distancia de Q_1 a la mediana es mayor que la distancia de la mediana a Q_3	Ambas distancias son iguales	La distancia de Q_1 a la mediana es menor que la distancia de la mediana a Q_3



Distribución asimétrica a la izquierda



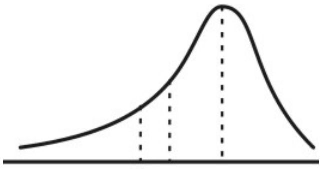
Distribución simétrica



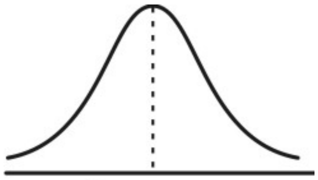
Distribución asimétrica a la derecha

5. MEDIDAS DE ASIMETRÍA Y CURTOSIS:

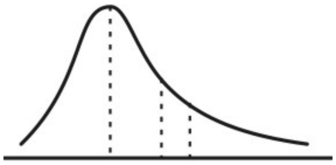
Los coeficientes de asimetría sirven para identificar y describir la tendencia de los datos a concentrarse o agruparse en alguna parte de la gráfica, en función a su frecuencia. Esta concentración de datos puede dar lugar a tres tipos de distribución:



Asimétrica negativa
(cola a la izquierda)



Simétrica



Asimétrica positiva
(cola a la derecha)

La fórmula de Pearson para el cálculo de la asimetría es:

$$AS = \frac{3\left(Media - Mediana\right)}{Desv.Estándar} = \frac{3\left(\bar{x} - Me\right)}{s}$$

Si AS = 0, entonces la distribución es simétrica.

Si AS > 0, entonces la distribucion es asimetrica positiva.

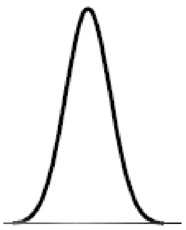
Si AS < 0, entonces la distribucion es asimetrica negativa.

Según Toma (2012) cuando en la distribución de datos de una población o muestra se encierra un igual porcentaje de datos en intervalos equidistantes por debajo y por encima de la media aritmética, entonces se dice que la distribución es simétrica. Pero si se encierran diferentes porcentajes de observaciones por debajo y por encima de la media, la distribución es asimétrica.

Como menciona Toma (2012) la curtosis mide es el grado de concentración de un conjunto de datos en relación con la media aritmética, es decir, mide la proximidad existente entre las observaciones y el promedio. En una distribución donde las observaciones son muy parecidas entre sí, se tendrá una gran proximidad con respecto al promedio y, por tanto, una alta concentración de datos o un alta agudeza; mientras que en una distribución donde las observaciones son muy diferentes entre sí, se tendrá que las observaciones estarán distantes del promedio y, por tanto, una baja concentración de los datos o una baja agudeza.

Platicúrtica
(Alta dispersión)

Mesocúrtica
(Moderada dispersión)



Leptocúrtica
(Baja dispersión)

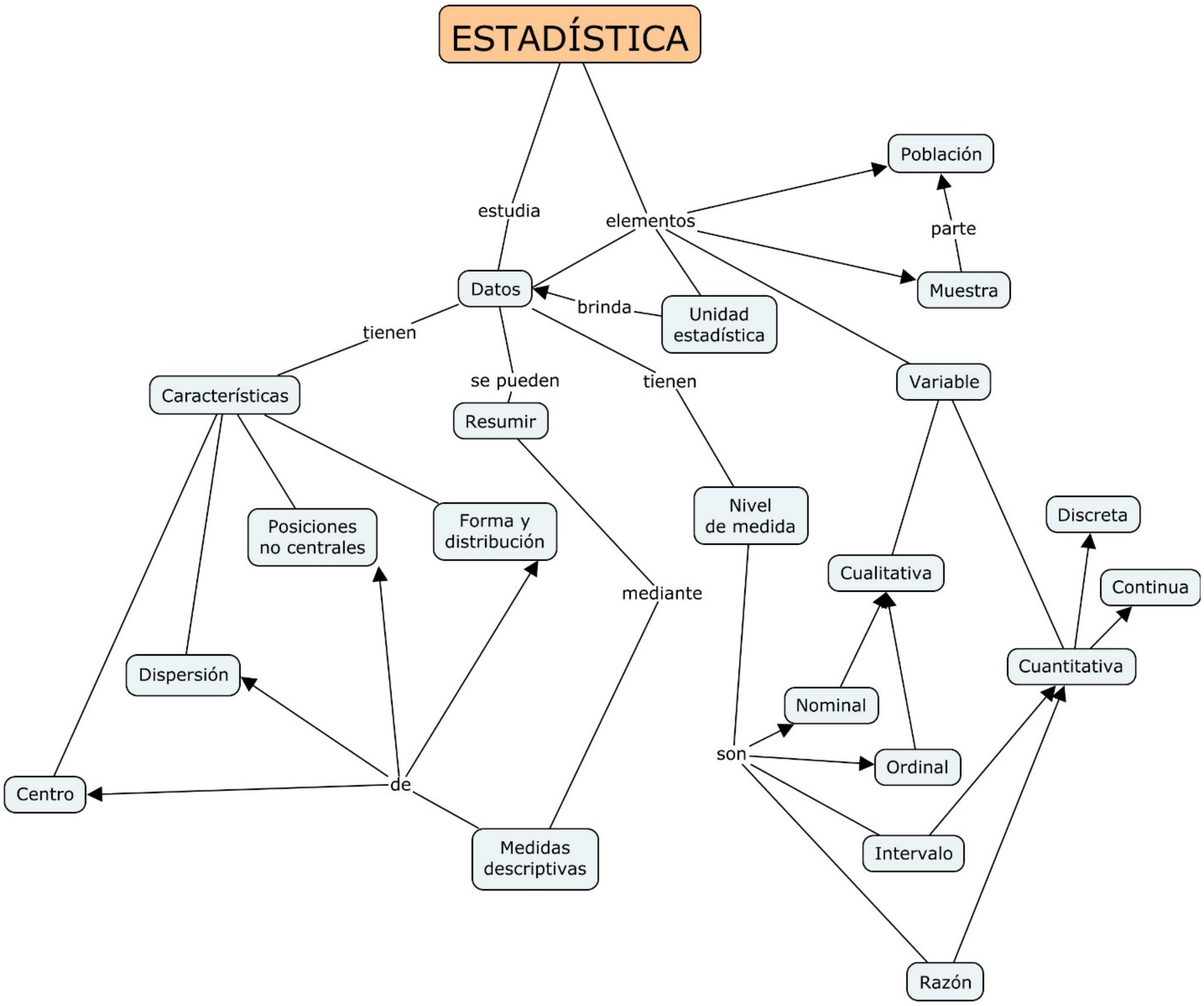
La fórmula para el cálculo de la curtosis es:

$$K = \frac{P_{75} - P_{25}}{2\left(P_{90} - P_{10}\right)}$$

Si K < 0,243 la distribución es platicúrtica.
Si 0,243 < K < 0,253 la distribución es mesocúrtica.
Si K > 0,253 la distribución es leptocúrtica.

Fuente: Jorge Toma Inafuko

EN SÍNTESIS:



Fuente: Elaboración propia

LECTURA SELECCIONADA N.º 1

¿Para qué sirve la mediana, si ya tenemos la media aritmética?

La media aritmética es una excelente medida de tendencia central. Sus buenas propiedades, junto con el hecho de ser fácil de entender y de calcular, la hacen muy usada y también muy apreciada (a veces demasiado, como cuando se pretende resumir solo en ella toda la información que contienen los datos), pero la mediana tiene unas propiedades de las que carece la media, por lo que es un buen complemento informativo e incluso en algunos casos puede ser una medida más útil. Estas propiedades son:

- Es más robusta que la media frente a la presencia de anomalías. Un ejemplo muy simple: supongamos que nuestros datos son: 2, 5, 6, 7 y 9. La media es 5,6 y la mediana es 6. Si al introducir los datos al ordenador nos equivocamos y en último lugar en vez de 9 introducimos 99, la media pasa a ser 23,8, mientras que la mediana sigue siendo 6.

En algunos casos, como cuando se trabaja con datos todavía no depurados, fijarse en la mediana puede ser más recomendable porque la información que da está menos afectada por las posibles anomalías que puedan existir.

- Por su propia definición, la mediana deja un 50% de las observaciones por encima y otro 50% por debajo y esto le da unas ventajas que la media no tiene. Por ejemplo, si queremos saber si en nuestra empresa estamos entre los que cobran más o entre los que cobran menos, debemos comparar nuestro salario con la mediana, y no con la media. Si sólo hay 10 trabajadores y los salarios son (pongamos que en miles de euros): 0,8; 0,8; 0,9; 0,9; 1,0; 1,0; 1,1; 1,1; 1,2 y 10, todos menos 1 (en este caso el 90%) están por debajo de la media, que es 1,88. Esto no pasa nunca con la mediana, si estamos por encima de la mediana, estamos con el 50% de los que más cobran

Otro ejemplo. Si un examen se aprueba sacando una nota igual o superior a 5 y la nota media que han sacado los estudiantes es de 5, no sabemos cuántos han aprobado. Si se han examinado 50 estudiantes, puede ser que 41 hayan suspendido con un 4; 8 estudiantes hayan sacado un 10 y uno haya obtenido un 6. Esto da media 5, aunque es verdad que son unas notas muy raras. Si la mediana es 5, seguro que la mitad han aprobado.

Además, puestos a criticar la media, podemos decir que su uso conduce a algunas situaciones paradójicas, como aquella que dice que la mayoría de los hombres tiene un número de piernas superior a la media.

Si la distribución de los datos es simétrica, la media y la mediana coinciden, y entonces todo son ventajas. Por ejemplo, en una distribución Normal la media y la mediana son iguales [$P(X>m) = 0,5$] y por tanto, si los valores que tenemos provienen de una Normal, y no hay anomalías, la media y la mediana no andarán muy lejos la una de la otra. En cualquier caso, si disponemos de un medio de cálculo fácil, podemos calcular las dos y aprovechar lo mejor de cada una.

LECTURA SELECCIONADA N.º 2

¿Cuándo conviene utilizar *boxplots* para analizar o describir datos?

Los *boxplots* son gráficos muy apropiados para mostrar el comportamiento de los datos cuando interesa presentarlos estratificados por alguna variable cualitativa.

Por ejemplo, la Figura 7.1 muestra la distribución de los pesos de 500 paquetes de azúcar llenados en una planta de envasado que consta de 5 líneas independientes. Los pesos se presentan según la línea en que se han llenado (100 paquetes por línea). El valor nominal es de 1.000 g.

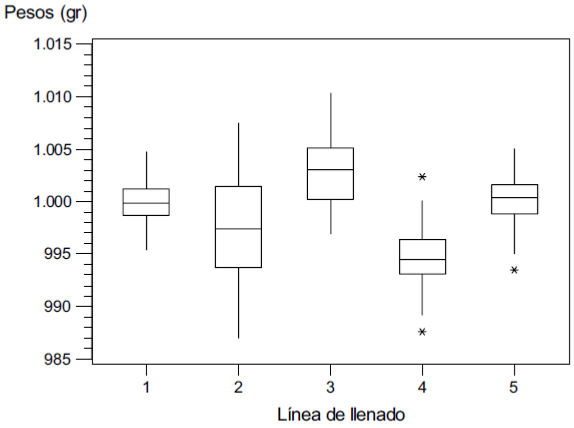


Figura 7.1. *Distribución de los pesos de paquetes de azúcar según la línea en que han sido llenados*

Fácilmente se puede observar que en las líneas 1 y 5 la producción está centrada en el valor objetivo con una variabilidad de aproximadamente ± 5 *g*, mientras que las otras líneas están descentradas, presentando la línea 2 una variabilidad claramente mayor que las otras.

Otra situación en la que el uso de *boxplots* resulta adecuado es la mostrada en la Figura 7.2, que representa los precios de venta de los pisos en Barcelona en 1990 (en miles de pesetas por metro cuadrado) según el distrito en que se encontraban1.

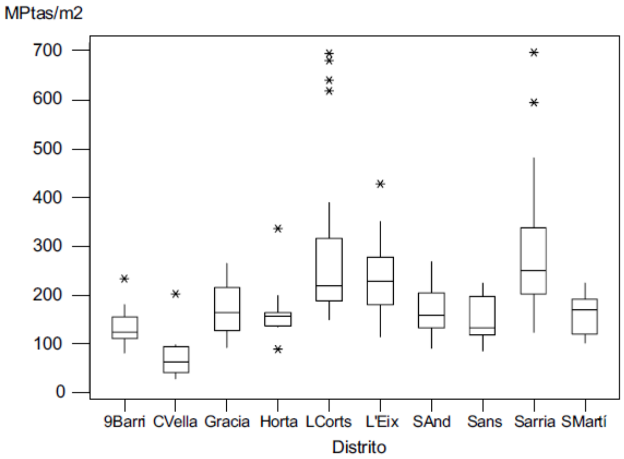


Figura 7.2. *Precio de venta de los pisos en Barcelona en 1990 (en miles de pesetas por metro cuadrado) según el distrito en que se encontraban*

En ambos casos los gráficos resumen de una forma clara y compacta la información que contienen los datos. Además facilitan la comparación entre grupos y permiten una rápida identificación de valores atípicos.

ACTIVIDAD FORMATIVA N.º 3

Resuelve los siguientes ejercicios e interpreta los estadígrafos que calcules. No olvides que la interpretación es muy importante en nuestra asignatura. De ser necesario vuelve a estudiar los ejemplos desarrollados.

INSTRUCCIONES:

Lea e interprete los diferentes contenidos y ejercicios presentados en los temas anteriores para que resuelva los ejercicios de la presente actividad:

1. Consideremos que un vendedor de cigarrillos vende de lunes a sábado el siguiente número de cajetillas.

   55 – 70 – 64 – 50 – 65 – 68 – 55

   - Calcule la media, mediana y moda del número de cajetillas vendidas. Interprete.
   - Calcule el rango, la varianza y la desviación estándar del número de cajetillas vendidas. Interprete.
   - Calcule el coeficiente de variación y determine si el número de cajetillas vendidas es un conjunto muy homogéneo, homogéneo, etc.

2. En una tarde de entrenamiento, se ha medido la estatura de los jugadores del equipo de fútbol de una universidad. Sus estaturas, en cm, se reflejan en la siguiente tabla agrupados en intervalos:

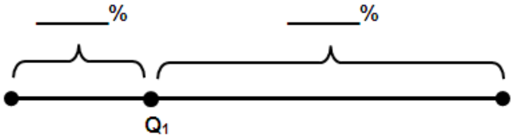
Alturas	Nº alumnos (fi)
[150,155>	3
[155,160>	7
[160,165>	6
[165,170>	4
[170,175>	5

   - Calcule las medidas de tendencia central (media, mediana, moda) e interprételas.
   - Calcule las medidas de dispersión absoluta (rango, varianza, desviación estándar) y las de dispersión relativa (coeficiente de variación). Interprételas.

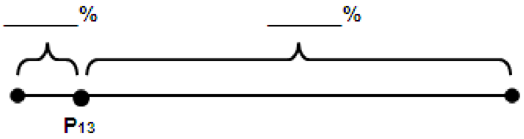
3. Un conjunto de 20 valores tiene una media aritmética igual a 50; con una desviación estándar de 3,6. Otro conjunto de 20 valores tienen una media aritmética igual a 30 con una desviación estándar de 4,2. Calcule los coeficientes de variación y luego determine cuál de ellos es más disperso.

4. Complete los gráficos con los porcentajes correspondientes a los cuartiles y percentiles que se indican:

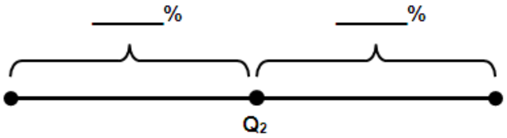
a)



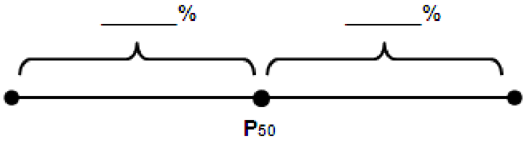
b)



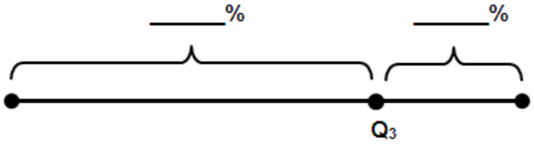
c)



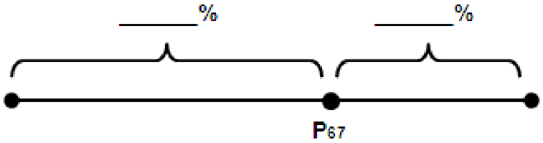
d)



e)



f)



5. Los datos que a continuación se muestran corresponden a las edades de 10 empleados de una planta de producción: 25, 32, 29, 48, 36, 35, 48, 30, 27, 20.

- Calcule el cuartil 1 y el cuartil 3. Interprételos.
- Calcule el percentil 90 y el percentil 10. Interprételos.
- Calcule el rango intercuartil. Interprételo.
- Calcule el coeficiente de asimetría. Interprételo.

6. En la granja "El Super Pollo", se registra la siguiente tabla de distribución de los pesos (en gramos) de los pollos beneficiados:

Peso	N° pollos
[1800 - 1900>	60
[1900 - 2000>	160
[2000 - 2100>	280
[2100 - 2200>	260
[2200 - 2300>	160
[2300 - 2400>	80
	1000

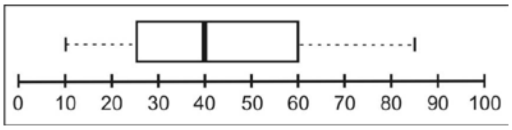
Para efectos de venta y marketeo la empresa los clasifica en tres categorías, de acuerdo a su peso:

- El 20% de los pollos menos pesados pertenecen a la categoría de "pollos tiernos".
- El 60% de los pollos menos pesados siguientes pertenecen a la categoría de "super pollos".
- El resto (20% restante), pertenecen a la categoría de "polli pavos"

Responda:

- ¿Cuáles son los límites de peso entre las categorías referidas?

7. Se da el siguiente diagrama de caja:



- ¿Cuál es la mediana, el valor mínimo y el máximo, el primer y tercer cuartil?
- ¿Estaría usted de acuerdo en que la distribución es simétrica? ¿Por qué? Explique.

8. Los pesos de un grupo de estudiantes del Centro Preuniversitario de la Universidad Nacional Mayor de San Marcos han dado los siguientes datos: Q1 = 50 kg; Q3 = 70 kg; Me = 55 kg .Elabore el diagrama de caja y bigote para dichos pesos. En el gráfico, diga usted ¿cómo se considerarían los pesos de 102 kg?

9. Una muestra de digitadoras de textos reveló que su rapidez media de tecleo es de 87 palabras por minuto, con una mediana de 73 palabras. La desviación estándar es de 16,9 palabras por minuto. ¿Cuál es el coeficiente de asimetría? Interprete la respuesta, luego conteste: ¿La mayoría de estas secretarias son rápidas tecleando o son lentas?

10. La siguiente tabla de distribución de frecuencias corresponde al rendimiento de gasolina en kilómetros por litro de una muestra de 50 automóviles producidos por Nissan México. Calcule el coeficiente de curtosis. Interprete.

RENDIMIENTO KM/L	N° DE AUTOS
[10.2-10.9>	3
[10.9-11.6>	8
[11.6- 12.3>	19
[12.3-13.0>	28
[13.0- 13.7>	19
[13.7- 14.4>	8
[14.4- 15.1>	3

Videos

Tema Nº 1

Media, Varianza y Desviación Estándar usando la calculadora CASIO fx-82MS para datos simples

https://www.youtube.com/watch?v=qguhqq0xvM0

Tema Nº 2

Media, Varianza y Desviación Estándar con tablas de frecuencia y calculadora CASIO fx-82MS

https://www.youtube.com/watch?v=2FPsN0oZsyU

Tema Nº 3

Media, Varianza y Desviación Estándar de datos simples usando la calculadora CASIO fx-82ES-Plus

https://www.youtube.com/watch?v=7s77eMDRkdc

Tema Nº 4

Calcular Media, Varianza y Desviación Estándar con calculadora CASIO fx-350ES

https://www.youtube.com/watch?v=vX0c3M1qZIs

ACTIVIDAD FORMATIVA N.º 4

Aplica la encuesta elaborada en la anterior unidad a una muestra de 30 personas. Procura seleccionar a los encuestados de manera aleatoria, ya que esto permitirá que tu muestra represente mejor a la población de donde se le está tomando.

INSTRUCCIONES:

- Antes de aplicar la encuesta primero realiza un monitoreo, para ello imprime 10 ejemplares y aplícalos a manera de prueba piloto. La finalidad de este procedimiento es poner a prueba el instrumento de recolección de datos en condiciones reales.
- Luego analiza las encuestas aplicadas buscando las posibles deficiencias que pudieran presentarse (instrucciones no comprensibles, preguntas no muy claras, preguntas que se prestan a más de una respuesta, etc.). Luego corrige la encuesta original tomando en cuenta las observaciones halladas en el monitoreo.
- Una vez corregida la encuesta, imprímela y reproduce la cantidad de encuestas a aplicar.
- Luego de aplicada la encuesta, traslada los datos a una matriz, en la que sólo copiarás las respuestas que dieron las personas que participaron de tu encuesta**.**

Ejemplo:

N°	PREG1	PREG2	PREG3	PREG4	PREG5	PREG6	PREG7	PREG8	PREG9	PREG10
1	23	FEMEN.	ADM. MARK.	NO	SNACKS	3	3.00	LLAMA ATENC.	MALO	RETIREN
2	22	FEMEN.	PSIC.	SÍ	GASEOSA	3	8.00	COMO PENSÉ	BUENO	SUFICIENTE
3	20	FEMEN.	ECON.	NO	GALLETAS	3	7.00	COMO PENSÉ	BUENO	RETIREN
4	23	MASC.	ADM. RR.HH.	SÍ	GASEOSA	3	5.00	LLAMA ATENC.	BUENO	PONGAN
5	18	MASC.	ADM. MARK.	NO	GOLOS.	2	3.00	PRECIO	MALO	SUFICIENTE
6	23	FEMEN.	CC.CC.	NO	GOLOS.	3	9.00	PRECIO	BUENO	SUFICIENTE
7	21	MASC.	ADM. NEG.	SÍ	CAFÉ	3	4.00	PRECIO	REGULAR	SUFICIENTE
8	19	FEMEN.	ECON.	NO	GOLOS.	2	7.00	PRECIO	MALO	SUFICIENTE
9	22	FEMEN.	ECON.	SÍ	QUEQUE	2	7.00	LLAMA ATENC.	MALO	SUFICIENTE
10	22	MASC.	CC.CC.	SÍ	CAFÉ	3	9.00	COMO PENSÉ	BUENO	RETIREN
11	24	MASC.	CC.CC.	NO	QUEQUE	1	9.00	LLAMA ATENC.	REGULAR	PONGAN
12	18	FEMEN.	ECON.	SÍ	GASEOSA	3	7.00	COMO PENSÉ	REGULAR	PONGAN
13	24	MASC.	CC.CC.	NO	SNACKS	2	9.00	PRECIO	REGULAR	SUFICIENTE
14	21	MASC.	ADM. FIN.	SÍ	GOLOS.	2	2.00	COMO PENSÉ	MALO	RETIREN
15	21	FEMEN.	ADM. FIN.	SÍ	QUEQUE	2	2.00	COMO PENSÉ	BUENO	SUFICIENTE
16	22	MASC.	ADM. MARK.	SÍ	GASEOSA	2	3.00	LLAMA ATENC.	BUENO	RETIREN
17	18	MASC.	CONTAB.	SÍ	GASEOSA	2	6.00	COMO PENSÉ	BUENO	PONGAN
18	23	FEMEN.	ADM. RR.HH.	NO	GASEOSA	2	5.00	LLAMA ATENC.	MALO	PONGAN
19	20	MASC.	ADM. MARK.	NO	GASEOSA	1	3.00	PRECIO	REGULAR	PONGAN
20	19	MASC.	ADM. NEG.	SÍ	QUEQUE	2	4.00	PRECIO	BUENO	RETIREN

- Finalmente, en el programa Excel, elabora y calcula para cada variable (es decir, cada pregunta):
  - Tablas de frecuencia de acuerdo al tipo de variable para organizar los datos.
  - Gráfico estadístico de acuerdo al tipo de variable para poder visualizar de manera más clara los resultados obtenidos.
  - Estadígrafos o medidas de resumen de acuerdo al tipo de variable y al nivel de medición. Interpreta (no olvides que esta parte del trabajo es la más importante).

PRUEBA DE DESARROLLO – UNIDAD II

1. Determine la veracidad (V) de las siguientes proposiciones:

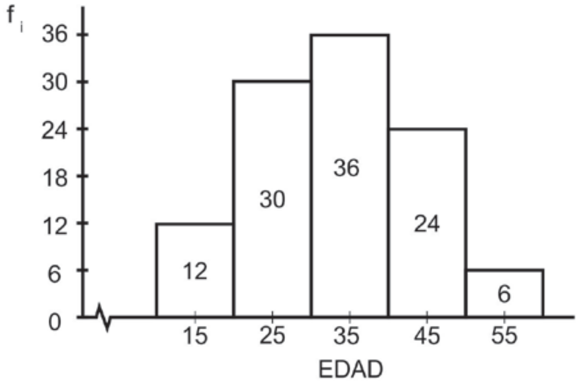
- La moda no siempre existe. Y si existe, no siempre es única ( )
- La mediana está afectada por los valores de los datos que se encuentran muy alejados ( )
- La media, mediana y moda de 3; 3; 3; 3; 3 siempre valen 3 ( )
- Si $\bar{x}$ = 22, Me = 23 y Mo = 25, entonces la distribución tiene cola a la derecha ( )
- La moda nunca se encuentra afectada por valores extremos ( )

2. Los pesos de una muestra de 45 de cajas enviadas como encomienda por la empresa "Cruz del Sur" se presentan resumidos en la siguiente tabla.

Pesos	f_i
[20 - 30>	4
[30 - 40>	8
[40 - 50>	15
[50 - 60>	12
[60 - 70>	6

- Calcule la media aritmética de los pesos. Interprete.
- Calcule la mediana de los pesos. Interprete.
- Luego de calcular el coeficiente de variación determine si la media aritmética es confiable. Explique por qué.

3. Calcule el coeficiente de asimetría del siguiente conjunto de edades.

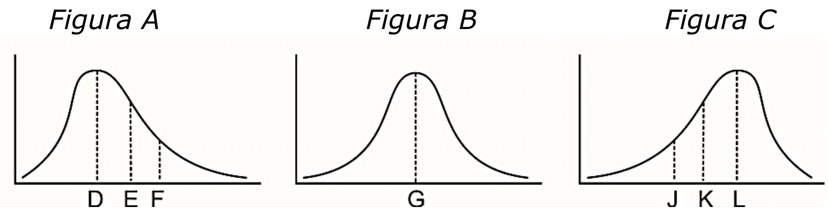


4. Se presenta a continuación el número de infracciones tributarias cometidas durante los meses de enero y febrero del presente año, por un grupo de empresas (los datos corresponden a muestras).

   - Calcule el coeficiente de variación, interprételo
   - Determine el mes en el que el número de infracciones cometidas fue más homogéneo.

ENERO	FEBRERO
6 ; 5 ; 9 ; 2 ; 4	$\bar{x}$ = 3 infracciones. Me = 4 infracciones Mo = 2 infracciones s2 = 1,44

5. Dadas las siguientes distribuciones de datos:



- ¿Qué medida de tendencia central representa la letra D? ¿La F?
- La moda en la figura B es $ 200. La mediana y la media entonces cuánto valen?
- En la figura C el valor de K es $ 100. Por lo tanto ¿la mitad de los valores está por encima de qué valor?
- ¿Por qué motivo la figura A presenta un sesgo hacia el lado derecho?

6. Un psicólogo ha tomado una muestra de 8 niños de tercer grado de primaria, con la finalidad de aplicar el test de la figura humana. Los tiempos (en minutos) que demoraron se muestran a continuación:

10 ; 14 ; 21 ; 18 ; 20 ; 17 ; 13 ; 19

Calcule el percentil 80 e interprételo.

GLOSARIO DE LA UNIDAD

A

Asimetría: Coeficiente que mide la nivel de dispersión de una serie de datos respecto a la media.

C

Cuartiles: Los tres valores que dividen datos ordenados en cuatro grupos, con aproximadamente el 25% de los valores en cada grupo.

D

Desviación Estandar: Medida de variación igual a la raíz cuadrada de la varianza.

M

Media: La suma de un conjunto de valores, dividida entre el número de valores.

Mediana: Valor que está a la mitad de un conjunto de valores acomodados en orden por magnitud.

Medida de tendencia central: Valor que pretende indicar el centro de los valores de una colección de datos.

Medida de variación: Cualquiera de varias medidas diseñadas para reflejar la magnitud de la variación o dispersión de un conjunto de valores.

Moda: Valor que se presenta con mayor frecuencia.

Multimodal: Se dice que un conjunto de datos es multimodal cuando tiene más de dos modas.

R

Rango de percentiles 10-90: Diferencia entre los percentiles décimo y nonagésimo.

S

Sesgado: No simétrico y que se extiende más hacia un lado que hacia el otro.

Sesgo negativo: Sesgado hacia la izquierda.

Sesgo positivo: Sesgado hacia la derecha.

V

Varianza: Promedio de las desviaciones cuadráticas de cada dato respecto a la media

BIBLIOGRAFÍA DE LA UNIDAD II

Devore, J. (2008). *Probabilidad y Estadística (*7 ed.). México: Editorial Cengage Learning Editores S.A.

Levine, D., Krehbiel, T., & Berenson, M. (2006). *Estadística para administración.* (4 ed.). Mexico: Editorial Pearson.

Martínez Bencardino, C. (2012). *Estadística y muestreo* (13 ed.). Bogotá, Colombia: Editorial Ecoe Ediciones.

Toma, J. (2012). *Estadística aplicada – Primera parte* (2 Ed.). Perú: Fondo editorial de la Universidad del Pacífico

Triola, M.(2009). *Estadística.* México: Pearson Educación.

AUTOEVALUACION No 2

1. El Departamento de control de calidad de la empresa Gloria S.A., verifica el peso de una muestra de frascos de mermelada de 125 g. Los pesos de la muestra de nueve frascos fabricados la hora anterior fueron:

   126,2 – 125,3 – 124,2 – 128,1 – 125,8 – 126,2 – 127,4 – 129,1 – 126,2

   - Calcule la media aritmética, la mediana y la moda de los pesos. Interprete estos estadígrafos.
   - Calcule el rango y la desviación estándar de los pesos. Interprete estos estadígrafos.
   - Si se compara los pesos de esta muestra con los pesos de otra muestra tomada el día anterior, cuya media = 125,6 g. y desviación estándar de 1,02 g. ¿Cuál de las muestras tiene sus datos más homogéneos?

2. Dada la distribución de frecuencias respecto a los alquileres de una muestra de familias que habitan el conjunto habitacional “Las Lomas”–La Perla – Callao.

Alquiler ($)	f_i
[25 - 30>	9
[30 - 35>	15
[35 - 40>	18
[40 - 45>	13
[45 - 50>	7

   - Calcule el percentil 64. Interprete.
   - ¿Cuánto pagan como máximo el 25% de las familias por alquiler?

3. La siguiente tabla presenta las velocidades (en km/h) de dos muestras de autos que pasaron por un punto de control de velocidad en la carretera central, los días 20 y 21 de setiembre. Determine en cuál de los dos días la media aritmética de las velocidades es más representativa.

Día 20

Velocidades en km/h						
60	68	50	76	82	69	71

Día 21

Velocidades	f_i
[30 - 40 >	3
[40 - 50 >	7
[50 - 60 >	20
[60 - 70 >	15
[70 - 80 >	9
[80 - 90 >	6

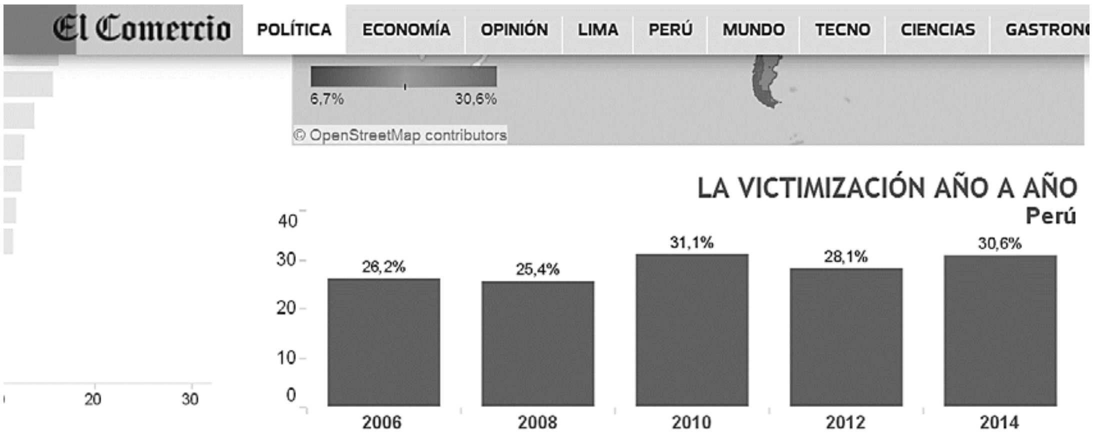
4. Se seleccionaron dos tipos distintos de alambres y en cada tipo se determinó la resistencia al peso (en kg) obteniéndose los siguientes resultados. Los datos son muestrales

Tipo 1 : 350 – 354 – 359 – 363 - 365 – 368 – 369 – 371 – 373 – 384 – 391 – 391 – 392
Tipo 2 : 350 – 362 – 361 – 359 – 375 – 382 – 371 – 365 – 368 – 370 – 385 – 394 – 395

Calcule el coeficiente de asimetría de Pearson y determine hacia dónde presentan asimetría cada una de los tipos de alambre. Interprete.

5. El día miércoles 22 de abril del presente año, el diario "El Comercio" publicó:

"No era una percepción producto de la histeria. El Perú es el país de América Latina con la mayor tasa de víctimas de la delincuencia. Así lo revela el Barómetro de las Américas 2014, presentado ayer en la sede del Instituto de Estudios Peruanos (IEP). De acuerdo a este trabajo del Proyecto de Opinión Pública de América Latina (LAPOP), el 30,6% de los ciudadanos peruanos ha sido víctima de al menos un acto delincuencial durante el año 2014."



Población en Perú el año 2014: 30 814 175 habitantes.

En relación a dicha publicación determine el valor de verdad (V) o falsedad (F) de las siguientes proposiciones:

- En el Perú, aproximadamente 21 385 037 habitantes no han sido víctimas de la delincuencia durante el año 2014 ( )
- El 68,9% de peruanos no fueron víctimas de la delincuencia el año 2010 ( )
- En el enunciado textual, el valor 30,6% es un parámetro ( )
- Aproximadamente el número de actos delictivos creció en 770354 del año 2012 al año 2014 ( )

UNIDAD III

"PROBABILIDAD"

DIAGRAMA DE PRESENTACIÓN DE LA UNIDAD III

RESULTADO DE APRENDIZAJE

Al finalizar la unidad, el estudiante será capaz de evaluar la probabilidad de un suceso y la probabilidad de ocurrencia en acontecimientos de sus actividades laborales.

CONTENIDOS	ACTIVIDADES FORMATIVAS (HABILIDADES Y ACTITUDES)	SISTEMA DE EVALUACIÓN (TÉCNICAS Y CRITERIOS)
TEMA N°1: PROBABILIDAD 1 Definiciones Básicas 2 Probabilidad 3 Enfoques de la probabilidad 4 Ley de los números grandes 5 Reglas del suceso infrecuente. **TEMA N° 2: REGLA DE LA SUMA Y REGLA DE LA MULTIPLICACIÓN** 1 Regla de adición 2 Regla de Multiplicación **TEMA N° 3: PROBABILIDAD TOTAL – TEOREMA DE BAYES** 1 Diagramas de árbol 2 Partición de un espacio Muestral 3 Probabilidad Total 4 Teorema de Bayes	· Identifica los elementos de los experimentos Aleatorios. · Calcula la probabilidad de eventos aleatorios y los interpreta. · Calcula la probabilidad de eventos aleatorios y los interpreta usando las reglas de adición y multiplicación. · Calcula las probabilidades condicionales utilizando el Teorema de Bayes. · Resuelve las actividades de la guía de práctica del texto, haciendo uso del Laboratorio de cómputo.	Procedimientos e indicadores de evaluación permanente · Entrega puntual de trabajos realizados. · Calidad, coherencia y pertinencia de contenidos desarrollados. · Prueba teórico-práctica, individual. · Actividades desarrolladas en sesiones tutorizadas. Criterios para evaluar habilidades para calcular e interpretar probabilidades · Calcula e interpreta las probabilidades de sucesos simples y compuestos. · Aplica correctamente las reglas de las operaciones con probabilidades a situaciones problemáticas. · Resuelve correctamente los ejercicios planteados.

RECURSOS:

Videos:

Tema N° 1

Historia de la probabilidad

https://www.youtube.com/watch?v=r_w3-uAnrZ8

Tema N° 2

Regla de adición

https://www.youtube.com/watch?v=3-zOxkssUoc

Regla de la multiplicación.

https://www.youtube.com/watch?v=ifTWwKH8AT0

Tema N° 3

Probabilidad condicional.

https://www.youtube.com/watch?v=Ik9NPdNgXhQ

Teorema de Bayes

https://www.youtube.com/watch?v=vX0c3M1qZls

Lectura complementaria:

Lectura seleccionada Nº 1
Historia y relevancia de la teoría de la probabilidad
Estadística para Administración y Economía.
Richard Levin. Pág. 128
Lectura seleccionada Nº 2
¿Debe preocuparse de que le realicen una prueba de detección de drogas cuando solicite un trabajo?
Estadística. Mario Triola. Pág. 137

INSTRUMENTO DE EVALUACIÓN	Prueba de desarrollo
BIBLIOGRAFÍA (BÁSICA Y COMPLEMENTARIA)	**BÁSICA** Devore, J. (2008). *Probabilidad y Estadística* (7 ed.). México: Editorial Cengage Learning Editores S.A. **COMPLEMENTARIA** Ross, S. (2001). *Probabilidad y Estadística para Ingenieros.* (3 ed.) México: Mc Graw Hill. Berenson, M. & Levine, D. (2010). *Estadística Básica en Administración, Conceptos y aplicaciones.* México: Prentice Hall.
RECURSOS EDUCATIVOS DIGITALES	· Instituto Nacional de Estadística e Informática. Disponible en: http://www.inei.gob.pe/ · Tokio explora en un congreso la “imprescindible” ciencia de la estadística. (2012, Jul 04). EFE News Service. Disponible en: http://search.proquest.com/docview/1023206430?accountid=146219 · Canales, E. (2005, Jul 26). Mexicar / AMLO sin estadistica. El Norte. Disponible en:: http://search.proquest.com/docview/311786309?accountid=146219

En alguna ocasión te has preguntado cuál es la probabilidad de que suceda algún suceso, como por ejemplo, la probabilidad de que mañana llueva, la probabilidad de que algún equipo de fútbol gane un partido, o la probabilidad de ganar un premio y hasta la probabilidad de acertar una pregunta con alternativas al marcarla al azar. Nos damos cuenta que vivimos en un mundo que es incapaz de predecir el futuro con total certeza lo cual genera el estudio y uso de la teoría de la probabilidad, la cual nos permitirá reconocer y ordenar nuestras suposiciones para poder tomar alguna decisión de manera más concreta y con fundamento. Te invito entonces a conocer el mundo de las probabilidades teniendo una probabilidad alta de que te va a gustar.

TEMA Nº 1: PROBABILIDAD

1. Definiciones básicas:

1.1. Experimento aleatorio:

Es cualquier acción o proceso cuyo resultado está sujeto a la incertidumbre. Ejemplo: Lanzar una moneda, lanzar un dado, pesar un trozo de pan, medir el tiempo que demora en resolver un ejercicio de matemática, etc.

1.2. Espacio muestral:

Es el conjunto de todos los posibles resultados de un experimento aleatorio. Se denota con la letra griega mayúscula "omega" (Ω).

Ejemplo:

Para el lanzamiento de un dado: Ω = { 1; 2; 3; 4; 5; 6 }

Para el peso de un trozo de pan: Ω = { x / x Î R Ù x > 0 }

1.3. Evento:

Es cualquier subconjunto de resultados contenidos en el espacio muestral. Los eventos se denotan con las primeras letras del alfabeto escritas en mayúscula.

Es decir:

A es un evento sí y sólo sí $A \subset \Omega$

A su vez, los eventos pueden ser:

- Simple: Si consiste en exactamente un resultado.

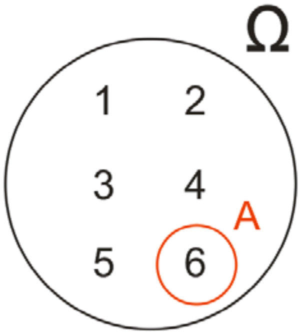
  Ejemplo:

  En el experimento de lanzar un dado:

  Espacio muestral Ω = { 1; 2; 3; 4; 5; 6 }

  Evento A = Número mayor que 5

  A = { 6 } es un evento simple



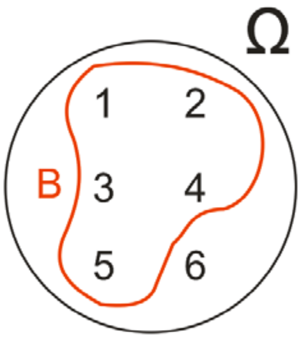
- Compuesto: Si consiste en más de un resultado.

  En el experimento de lanzar un dado:

  Espacio muestral Ω = { 1; 2; 3; 4; 5; 6 }

  Evento B = Número menor que 6

  B = { 1; 2; 3; 4; 5 } es un evento compuesto



1.4. Complemento de un evento:

El complemento de un evento A, denotado por A', es el conjunto de todos los resultados en Ω que no están contenidos en A.

A′ es el complemento del evento A sí y sólo sí $A' = \{x \in \Omega \wedge x \notin A\}$

Ejemplo:

- Evento A = Número mayor que 5

  A = { 6 }

  Complemento de A: A' = A^c = { 1; 2; 3; 4; 5 }

- Evento B = Número menor que 6

  B = { 1; 2; 3; 4; 5 }

  Complemento de B: B′ = B^c = { 6 }

1.5. Evento seguro:

Es aquel evento que de todas maneras tiene que ocurrir al efectuar el experimento. Es decir:

A es un evento seguro sí y sólo si $P(A)=1$

Ejemplo:

- En el experimento de lanzar un dado: El evento consistente en obtener un número mayor que 0 es un evento seguro, ya que de todas maneras va a ocurrir cuando se efectúe el experimento.

1.6. Evento imposible:

Es aquel evento que nunca va a ocurrir al efectuar el experimento. Es decir

A es un evento imposible sí y sólo si P(A)=0

Ejemplo:

- En el experimento de lanzar un dado: El evento consistente en obtener un número mayor que 7 es un evento imposible, ya que nunca va a ocurrir al efectuar el experimento, debido a que no es un posible resultado.

2. Probabilidad:

Como menciona Devore (2008), el objetivo de la probabilidad es asignar a cada evento A un número P(A), llamado probabilidad del evento A, el cual dará una medida precisa de la posibilidad de que A ocurra.

Además: $0 \leq P(A) \leq 1$

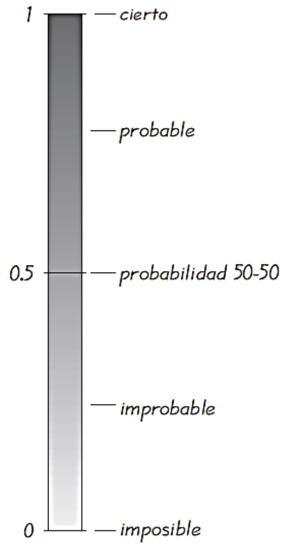


Figura 13. Valores posibles para probabilidades
Fuente: Mario Triola

3. ENFOQUES DE LA PROBABILIDAD:

3.1. Probabilidad clásica o a priori:

No es necesario efectuar el experimento. Se basa en la consideración de que los resultados de un experimento son igualmente posibles. Se calcula de la siguiente manera:

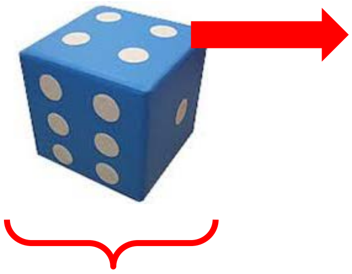
$$P(A) = \frac{Número\ de\ resultados\ \ favorables}{Número\ total\ \ de\ resultados\ \ posibles}$$

Ejemplo:

El experimento consiste en observar la cara que muestre hacia arriba al caer un dado. ¿Cuál es la probabilidad de que caiga un "cuatro"?

Solución:

Sea el evento A = Obtener un "cuatro"



Un resultado favorable: 4

Seis posibles resultados: 1; 2; 3; 4; 5; 6

$$P(A)=\frac{Número\ de\ resultados\ \ favorables}{Número\ total\ de\ resultados\ \ posibles}=\frac{1}{6}=0{,}167$$

3.2. Probabilidad por frecuencia relativa o a posteriori

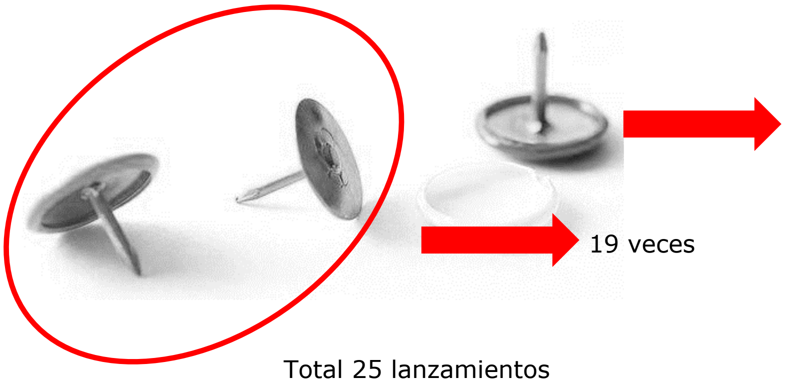
Es necesario efectuar el experimento. Se determina observando la proporción de veces que ocurrió en el pasado el evento. Se calcula de la siguiente manera:

$$\boxed{P(A)=\frac{Número\ de\ veces\ que\ el\ evento\ ocurrió\ en\ el\ \ pasado}{Número\ total\ de\ observaciones}}$$

Ejemplo:

Se quiere determinar la probabilidad de que al lanzar un chinche caiga con la punta hacia arriba. Para ello se ha lanzado el chinche 25 veces y en 6 de ellas cayó con la punta hacia arriba.

Solución:



6 veces

19 veces

Total 25 lanzamientos

$$P(A)=\frac{\text{Número de veces que el evento ocurrió en el pasado}}{\text{Número total de observaciones}}=\frac{6}{25}=0,24$$

3.3. Probabilidad subjetiva:

Si existe poca o ninguna experiencia en la cual se pueda basar una probabilidad entonces ésta puede obtenerse en forma subjetiva, es decir, asignada por una persona con base en cualquier información de que disponga. El valor de probabilidad estimado varía de una persona a otra.

Ejemplo:

¿Cuál es la probabilidad de que llueva en Huancayo este fin de semana?

Solución:

Al no haber mayor información podría una persona estimar que la probabilidad es de 0,3; en tanto que otra podría decir que es de 0,5. ¿Sería mejor solicitar esta estimación a un meteorólogo?

4. LEY DE LOS NÚMEROS GRANDES:

Según Triola (2009): Conforme un procedimiento se repite una y otra vez, la probabilidad de frecuencias relativas de un suceso, tiende a aproximarse a la probabilidad real.

Es decir, cuantas más veces se repite un experimento más se acerca su probabilidad al verdadero valor.

5. REGLA DEL SUCESO INFRECUENTE:

Conforme a Triola (2009): Si, bajo un supuesto dado, la probabilidad de un suceso particular observado es extremadamente pequeña, concluimos que el supuesto probablemente es incorrecto.

Es decir, se rechazan las explicaciones basadas en probabilidades muy bajas.

ACTIVIDAD FORMATIVA N.º 1

Interpreta y calcula las probabilidades de determinados eventos.

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

Probabilidades

1. Identificación de valores de probabilidad: Exprese el grado indicado de probabilidad como un valor numérico admisible de probabilidad (un número decimal comprendido en 0 y 1, inclusive).

    a. "Usted tiene una probabilidad de 50 – 50 de escoger el camino correcto".

    b. "Hay un 20% de probabilidad de que llueva mañana"

    c. "Usted tiene una probabilidad de un pelo de rana de casarse con mi hija".

    d. "Hay un 95% de probabilidad que mañana llueva"

    e. "Definitivamente, por la noche oscurecerá".

    f. "Usted tiene una probabilidad de 1 en 10 de aprobar el curso".

2. ¿Cuáles de los siguientes valores no pueden ser probabilidades?

    0 ; 1 ; -1 ; 2 ; 0,0123 ; 3/5 ; 5/3 ; $\sqrt{2}$

    3. Indique la probabilidad que se pide:

    a. ¿Cuál es la probabilidad de un suceso imposible?

    b. ¿Cuál es la probabilidad de un suceso inevitable o seguro?

    c. Un espacio muestral consiste en 10 sucesos separados que son igualmente probables. ¿Cuál es la probabilidad de cada uno?

    d. En un examen de verdadero / falso. ¿Cuál es la probabilidad de responder una pregunta correctamente si usted elige al azar?

    e. En un examen de opción múltiple, con cinco posibles respuestas para cada pregunta. ¿Cuál es la probabilidad de responder una pregunta correctamente si usted elige al azar?

4. Se lanza una moneda tres veces. Calcular la probabilidad de obtener por lo menos 1 sello.

5. En una caja hay 20 tarjetas numeradas del 1 al 20. Se extrae una tarjeta al azar. ¿Cuál es la probabilidad de que la tarjeta extraída contenga un número múltiplo de 2 o múltiplo de 3?

6. En una caja hay 6 bolas rojas, 3 blancas y 2 negras. Se extrae al azar una bola, ¿cuál es la probabilidad de que la bola extraída?

    - Sea roja
    - Sea blanca

- Sea negra
- No sea negra

7. Determine el valor de la probabilidad del complemento de A, si P(A) = 0.274

8. En una urna hay bolas rojas, blancas y verdes. Si la probabilidad de extraer una bola roja es 0,8. ¿Cuál es la probabilidad de extraer una bola que no sea roja?

9. En los siguientes ejercicios considere un suceso como "infrecuente" si su probabilidad es igual o menor que 0,05.

   a. En una encuesta del MINSA se interrogó a 1038 adultos acerca de los efectos del tabaquismo pasivo, 52 de ellos indicaron que tales efectos "no son dañinos en absoluto".

      - Si usted selecciona al azar a uno de los adultos que se encuestaron, ¿cuál es la probabilidad de seleccionar a alguien que opine que ser fumador pasivo no es dañino?
      - ¿Es infrecuente que alguien opine que ser fumador pasivo no es dañino en absoluto?

   b. En su primera cita, Katy le pide a Luis que adivine su fecha de nacimiento, sin importar el año.

      - ¿Cuál es la probabilidad de que Luis adivine correctamente (ignore los años bisiestos).
      - ¿Sería infrecuente que él adivinara con acierto en el primer intento?
      - Si usted fuera Katy, y Luis adivinara correctamente en su primer intento, ¿creería que él tuvo un golpe de suerte o estaría convencida de que él ya sabía la fecha en que usted nació?

10. Sea el experimento: "Se lanzan dos monedas simultáneamente".

    - Elabore el espacio muestral mediante un diagrama de árbol.
    - ¿Son los resultados mutuamente excluyentes?
    - Calcule la probabilidad de obtener 2 caras.
    - Calcule la probabilidad de que la primera moneda arroje cara.
    - Calcule la probabilidad de obtener un sello y una cara.
    - Calcule la probabilidad de no obtener ninguna cara.

VIDEOS

Historia de la probabilidad

https://www.youtube.com/watch?v=r_w3-uAnrZ8

TEMA Nº 2:

REGLA DE LA SUMA Y REGLA DE LA MULTIPLICACIÓN REGLAS BÁSICAS PARA EL CÁLCULO DE PROBABILIDADES

Permiten calcular probabilidades basadas en la combinación de eventos. Las dos reglas básicas que estudiaremos son:

1. Reglas de adición

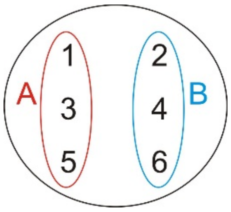
Regla especial de la adición

Eventos mutuamente excluyentes: Dos o más eventos son mutuamente excluyentes si no pueden ocurrir al mismo tiempo. Es decir:

> Dos eventos A y B son mutuamente excluyentes sí y sólo sí $A \cap B = \varnothing$

Ejemplo: En el experimento de lanzar un dado, se denotan dos eventos: A = Número impar, B = Número par. ¿Son estos dos eventos mutuamente excluyentes?

Solución:

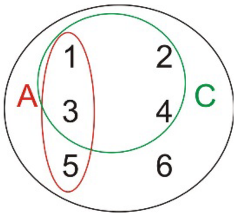


$A = \{1; 3; 5\}$
$B = \{2; 4; 6\}$
$A \cap B = \varnothing$

Observando se puede apreciar que los eventos A y B tienen una intersección nula o vacía. Es decir, no hay número alguno que sea a la vez par e impar al mismo tiempo. Por lo tanto, se puede decir que A y B sí son eventos mutuamente excluyentes.

Ejemplo: En el experimento de lanzar un dado, se denotan dos eventos: A = Número impar, C = Número menor que 5. ¿Son estos dos eventos mutuamente excluyentes?

Solución:



$A = \{1; 3; 5\}$
$B = \{1; 2; 3; 4\}$
$A \cap B \neq \varnothing$

Observando el gráfico se puede apreciar que los eventos A y C sí tienen intersección diferente del vacío. Es decir, hay números (1 y 3) que son a la vez impares y menores que 5. Por lo tanto se puede decir que A y C no son eventos mutuamente excluyentes.

Para aplicar la regla especial de la adición los eventos deben ser mutuamente excluyentes. "Si dos eventos A y B son mutuamente excluyentes, la regla especial de la adición indica que la probabilidad de que ocurra uno u otro de los eventos es igual a la suma de sus probabilidades". Simbólicamente:

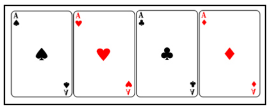
$$\boxed{A \cap B = \varnothing \rightarrow P(A\,o\,B) = P(A) + P(B)}$$

Ejemplo:

Sea A el evento de sacar un As de una baraja estándar de 52 cartas y B sacar un Rey de corazón rojo. Calcular la probabilidad de sacar un As o un Rey de corazón rojo en una sola extracción.

Solución:

Observando la baraja se cuenta en total 52 cartas. De ellas:

Evento A = {  }

P(A) = 4/52

Evento B = {  }

P(B) = 1/52

Por lo visto, A y B son eventos mutuamente excluyentes ya que los eventos tienen intersección nula.

Aplicando la regla especial de la adición

P(A o B) = P(A) + P(B)

= 4/52 + 1/52 = 5/52

= 0,0962

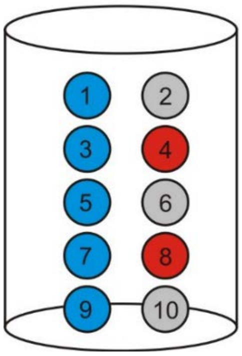
En una urna existe 10 bolas numeradas del 1 al 10. ¿Qué probabilidad existe de sacar en una sola extracción una bola enumerada con un número impar o con un número múltiplo de 4?

Solución:

Sean los eventos D = Bola con número impar

E = Bola con número múltiplo de 4

Observando se tiene:



D = {1; 3; 5; 7; 9}
P(D) = 5/10

E = {4; 8}
P(E) = 2/10

$D \cap E = \varnothing$

Son eventos mutuamente excluyentes
Aplicando la regla especial de la adición

P(D o E) = P(D) + P(E)
= 5/10 + 2/10 = 7/10
= 0,7

Regla general de la adición

Los resultados de un experimento pueden no ser mutuamente excluyentes, es decir, presentan intersección o se dice que se traslapan.

Cuando dos eventos se traslapan, a la probabilidad se le denomina probabilidad conjunta, la cual mide la posibilidad de que dos o más eventos ocurran en forma simultánea.

Para aplicar la regla general de la adición los eventos no deben ser mutuamente excluyentes. "Si dos eventos A y B no son mutuamente excluyentes, la regla general de la adición indica que la probabilidad de que ocurra uno u otro de los eventos es igual a la suma de sus probabilidades menos la probabilidad conjunta de dichos eventos". Simbólicamente:

$$\boxed{P(A\,o\,B)=P(A)+P(B)-P(A\,y\,B)}$$

Nota 1: Se resta P(A y B) porque se está contando dos veces esta probabilidad conjunta.

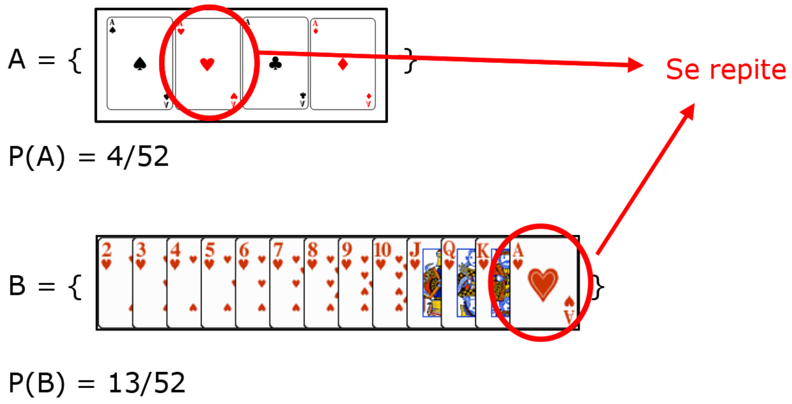
Nota 2: La fórmula $P\left(A\,o\,B\right)=P\left(A\right)+P\left(B\right)-P\left(A\,y\,B\right)$ incluye al caso en que los eventos A y B son mutuamente excluyente, pues cuando los eventos A y B son mutuamente excluyentes P(A y B) = 0

Ejemplo:

Sea A el evento de sacar un As de una baraja estándar de 52 cartas y B sacar un corazón rojo. Calcular la probabilidad de sacar un As o un corazón rojo en una sola extracción.

Solución:

Observando en la baraja se cuenta un total de 52 cartas, de ellas:



Podemos ver que la carta As de corazón se repite en ambos eventos. Es decir, no son mutuamente excluyentes.

Probabilidad conjunta: P(A y B) = 1/52

Aplicando la regla general de la adición:

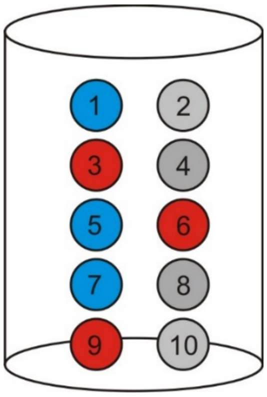
P(A o B) = P(A) + P(B) – P(A y B)
= 4/52 + 13/52 – 1/52
= 16/52
= 0,308

En una urna existe 10 bolas numeradas del 1 al 10. ¿Qué probabilidad existe de sacar en una sola extracción una bola enumerada con un número impar o con un número múltiplo de 3?

Solución:

Sean los eventos D = Bola con número impar
E = Bola con número múltiplo de 3

Observando se tiene:



D = {1; 3; 5; 7; 9}
P(D) = 5/10

E = {3; 6; 9}
P(E) = 3/10

No son eventos mutuamente excluyentes ya que dos números, el 3 y el 9, se repiten en ambos eventos.

Probabilidad conjunta: P(D y E) = 2/10

Aplicando la regla general de la adición

P(D o E) = P(D) + P(E) – P(D y E)
= 5/10 + 3/10 – 2/10 = 6/10
= 0,6

2. Reglas de multiplicación

Regla especial de la multiplicación:

Eventos independientes: Dos eventos son independientes si la ocurrencia de uno de ellos no tiene efecto en la probabilidad de la ocurrencia del otro. Caso contrario se dice que son dependientes.

Ejemplo:

Lanzar al aire dos veces una moneda son eventos independientes por que el resultado del primer evento no afecta sobre las probabilidades efectivas de que ocurra cara o sello, en el segundo lanzamiento.

Ejemplo:

Extraer focos defectuosos de una caja que contiene 10 focos buenos y 3 defectuosos. Si las extracciones se hacen con reposición (es decir, se devuelve el foco que se sacó primero antes de hacer la siguiente extracción) los eventos son independientes, ya que la cantidad de focos sigue siendo la misma y por lo tanto las probabilidades no se ven afectadas.

Ejemplo:

Extraer focos defectuosos de una caja que contiene 10 focos buenos y 3 defectuosos. Si las extracciones se hacen sin reposición (es decir, no se devuelve el foco que se sacó primero y se hace la siguiente extracción) los eventos son dependientes, ya que la cantidad de focos no sigue siendo la misma y por lo tanto las probabilidades sí se ven afectadas.

Para aplicar la regla especial de la multiplicación se requiere que los eventos sean independientes. Supone que un segundo resultado no depende del primero. "Si hay dos eventos independientes A y B, la probabilidad de que ocurran A y B se obtiene al multiplicar las dos probabilidades". Simbólicamente

$$\boxed{P(A\,y\,B)=P(A)\,.\,P(B)}$$

Ejemplo:

Se lanzan al aire dos monedas. ¿Cuál es la probabilidad de que ambas caigan mostrando "cara"?

Solución:

Al analizar el experimento podemos apreciar que lo que ocurra en la primera moneda no afecta en nada al resultado de la segunda moneda.

Sean los eventos:

A = Cara en la primera moneda
P(A) = 1/2

B = Cara en la segunda moneda
P(B) = 1/2

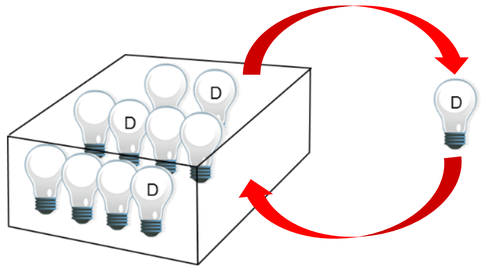
Podemos afirmar que los eventos A y B son independientes.
Aplicando la regla especial de la multiplicación:

$$\begin{aligned} P(A \text{ y } B) &= P(A)\,.\,P(B) \\ &= 1/2\,.\,1/2 = 1/4 \\ &= 0{,}25 \end{aligned}$$

Suponga que hay 10 focos en una caja, y se sabe que 3 están defectuosos. Se extrae primero uno de los focos y antes de extraer el segundo, se devuelve el primero a la caja. Calcule la probabilidad de que ambos sean defectuosos.

Solución

Si se extrae el primer foco defectuoso y se devuelve a la caja, entonces la probabilidad de extraer otro foco defectuoso es la misma que la del primero.



Si se repone el foco extraído sigue
habiendo la misma cantidad
de focos buenos y defectuosos.

Sean los eventos:

A = Primer foco defectuoso
P(A) = 3/10

B = Segundo foco defectuoso
P(B) = 3/10

Por lo tanto se dice que los eventos son independientes.
Aplicando la regla especial de multiplicación:

$$\begin{aligned} P(A \text{ y } B) &= P(A) \,.\, P(B) \\ &= 3/10 \,.\, 3/10 = 9/10 \\ &= 0{,}9 \end{aligned}$$

Regla general de la multiplicación:

Se utiliza para determinar la probabilidad conjunta de que ocurran dos eventos. En general, la regla indica que para dos eventos A y B, la probabilidad conjunta de que ambos eventos sucedan, resulta al multiplicar la probabilidad de que el evento A suceda, por la probabilidad condicional de que ocurra el evento B. De manera simbólica:

$$P(A \, y \, B) = P(A) \,.\, P(B \mid A)$$

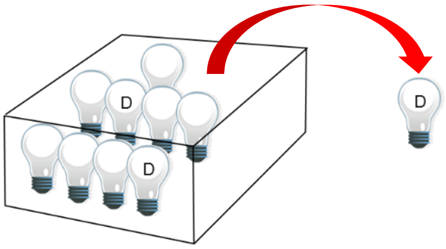
Donde $P(B \mid A)$ expresa la probabilidad condicional de que ocurra B dado que ya ocurrió A. La línea vertical "|" significa "Dado que".

Ejemplo:

Suponga que hay 10 focos en una caja, y se sabe que 3 están defectuosos. Se extrae primero uno de los focos, después se extrae un segundo foco de la caja sin restituir el primer foco (es decir, sin reposición). Calcule la probabilidad de que ambos sean defectuosos.

Solución:

Si se extrae el primer foco defectuoso y no se devuelve a la caja, entonces la probabilidad de extraer otro foco defectuoso no es la misma que la del primero, debido a que ya no se cuenta el foco extraído anteriormente.



Si no se repone el foco extraído ya no hay misma cantidad de focos buenos y defectuosos.

Sean los eventos:
A = Primer foco defectuoso
P(A) = 3/10

Luego de la extracción del primer foco:
B = Segundo foco defectuoso
P(B) = 2/9

Por lo tanto se dice que los eventos son independientes.
Aplicando la regla especial de multiplicación:
P(A y B) = P(A) . P(B)
= 3/10 . 2/9 = 6/90
= 0,0667

Nota: Si dos eventos A y B son independientes, aplicando la regla general de la multiplicación se tiene:

$$P(A\,y\,B)=P(A)\,\boxed{P(B\mid A)}=P(A)\,\boxed{P(B)}$$

Ya que el evento A no afecta la probabilidad de la ocurrencia de B.

ACTIVIDAD FORMATIVA Nº 2

Utiliza las reglas de las operaciones con probabilidades.

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

Regla de adición:

1. En cierta ciudad, la probabilidad de que una familia tenga televisor es 0,8; que tenga lavadora es de 0,6 y de que tengan ambos electrodomésticos es 0,45. ¿Cuál es la probabilidad de que una familia tenga televisor o lavadora o ambas cosas?
2. En la Facultad de Ciencias de la Empresa de la Universidad Continental, el 25% de los estudiantes desaprobó Matemática Básica, el 15% desaprobó Economía y 10% desaprobaron las dos asignaturas. Se selecciona al azar a un estudiante, ¿Qué probabilidad hay de que desaprobara Matemática Básica o Economía?
3. De una baraja de naipes de 52 cartas, ¿cuál es la probabilidad de que al extraer una sola carta, ésta sea un as de espadas o un dos de diamantes?
4. Una caja contiene 200 piezas, de las cuales 100 son producidas por la máquina A, 60 son producidas por la máquina B y el resto por la máquina C. Si una pieza es escogida al azar de la caja, ¿cuál es la probabilidad de que fuera producida por la máquina A o por la máquina C?
5. ¿Cuál es la probabilidad de que al lanzar dos dados aparezca la suma 7 o 10?
6. De 150 pacientes examinados en una clínica se encontró que 90 tenían enfermedades cardiacas, 50 tenían diabetes y 30 tenían ambos padecimientos. ¿Qué probabilidad hay de que si se elige un paciente al azar padezca de alguna de las enfermedades?
7. Se examinaron las tarjetas de registro de 200 empleados bancarios, en relación a ciertos idiomas. Se encontró que 100 de ellos hablaban inglés, 80 hablaban francés y 60 ambos idiomas. Si de este grupo se elige a un empleado, ¿cuál es la probabilidad de que hable inglés o francés o ambos idiomas?
8. En una agencia bancaria, el 5% de empleados eran estudiantes de maestría, el 10% eran estudiantes de diplomados y el 2% estudiaban maestría y diplomado a la vez. Si se selecciona al azar a uno de los empleados, ¿cuál es la probabilidad de que sea un estudiante de maestría o de diplomado o de ambas cosas a la vez?

Regla de multiplicación:

9. De una urna que contiene 5 bolas negras y 3 blancas, se extraen 3 de ellas en sucesión y sin reemplazo. ¿Cuál es la probabilidad de que las tres sean negras?
10. De una baraja ordinaria de cartas se extraen cinco de ellas sin reemplazo. ¿Cuál es la probabilidad de que las cinco sean corazones?
11. Tenemos que aplicar cuatro pruebas de inteligencia, tres de personalidad y dos de vocaciones profesionales. Si aplicamos estas nueve pruebas aleatoriamente, ¿cuál es la probabilidad de que sean aplicadas en primer lugar las cuatro pruebas de inteligencia y en segundo lugar las tres de personalidad y en último lugar las dos de intereses profesionales?

12. Tenemos cinco tarjetas marcadas respectivamente con las letras A, B, C, D y E. Las barajamos perfectamente y las vamos descubriendo una tras otra (sin volver la carta una vez descubierta). ¿Cuál es la probabilidad de que aparezcan según el orden D, E, C, A, B?

13. Si una caja contiene cuatro bolas blancas y cinco negras, ¿cuál es la probabilidad de que la primera bola extraída sea blanca y la segunda negra, si es que se cumple con devolver o reponer a la caja la primera bola extraída?

14. En una ciudad grande, el 70% de los hogares compra un periódico matutino y el 90% un periódico vespertino. Suponiendo que estos dos eventos son independientes, ¿qué probabilidad hay de que un hogar escogido al azar sea uno de los que compra ambos periódicos?

15. Un hombre posee un negocio y es, además, propietario de su casa. En un año cualquiera, la probabilidad de que la casa sea robada es 0,08 y la probabilidad de que su negocio sea robado es 0,14. Suponiendo que estos eventos sean independientes, calcule la probabilidad de que:

    - Sufra robos en ambos lugares en este año.
    - No se presenten robos en ninguno de los dos.

16. Se extraen 3 cartas sin reposición de una baraja de 52 cartas. ¿Cuál es la probabilidad de que: a) En la primera carta aparezca K y en las siguientes no aparezca, b) En la tercera carta aparezca la primera K?

17. En un solo lanzamiento de dos dados, ¿qué probabilidad tenemos de sacar dos cuatros?

18. Supongamos que un taller dispone de dos máquinas. En la primera se produce el 1,5% de unidades defectuosas y en la segunda el 3%. ¿Cuál es la probabilidad de que al extraer una de cada máquina, las dos sean defectuosas?

19. Se lanzan 3 monedas, ¿cuál es la probabilidad de que las tres sean caras?

20. Cuatro amigos se dirigen a un lugar, toman 4 rutas diferentes de acuerdo al riesgo que se corre de tener algún accidente. Si se le asignan las probabilidades de riesgo para cada ruta: 0,2; 0,15; 0,25; 0,1. Encuentre la probabilidad:

- De que ninguno sufra accidentes.
- Que los cuatro sufran accidentes.
- Que los dos primeros sufran accidentes y los restantes no.

VIDEOS.

Regla de adición

https://www.youtube.com/watch?v=3-zOxkssUoc

Regla de la multiplicación.

https://www.youtube.com/watch?v=ifTWwKH8AT0

TEMA Nº 3:
PROBABILIDAD TOTAL – TEOREMA DE BAYES

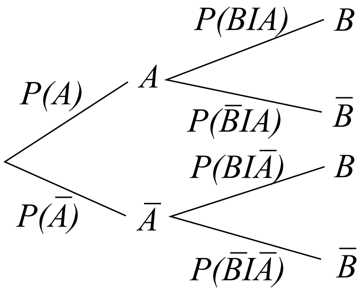
1. Diagramas de árbol

Un diagrama de árbol es muy útil para representar probabilidades condicionales y conjuntas.

Para la construcción de un diagrama en árbol se partirá poniendo una rama para cada una de las posibilidades, acompañada de su probabilidad.

En el final de cada rama parcial se constituye a su vez, un nudo del cual parten nuevas ramas, según las posibilidades del siguiente paso, salvo si el nudo representa un posible final del experimento (nudo final).

Hay que tener en cuenta que la suma de probabilidades de las ramas de cada nudo debe ser 1.



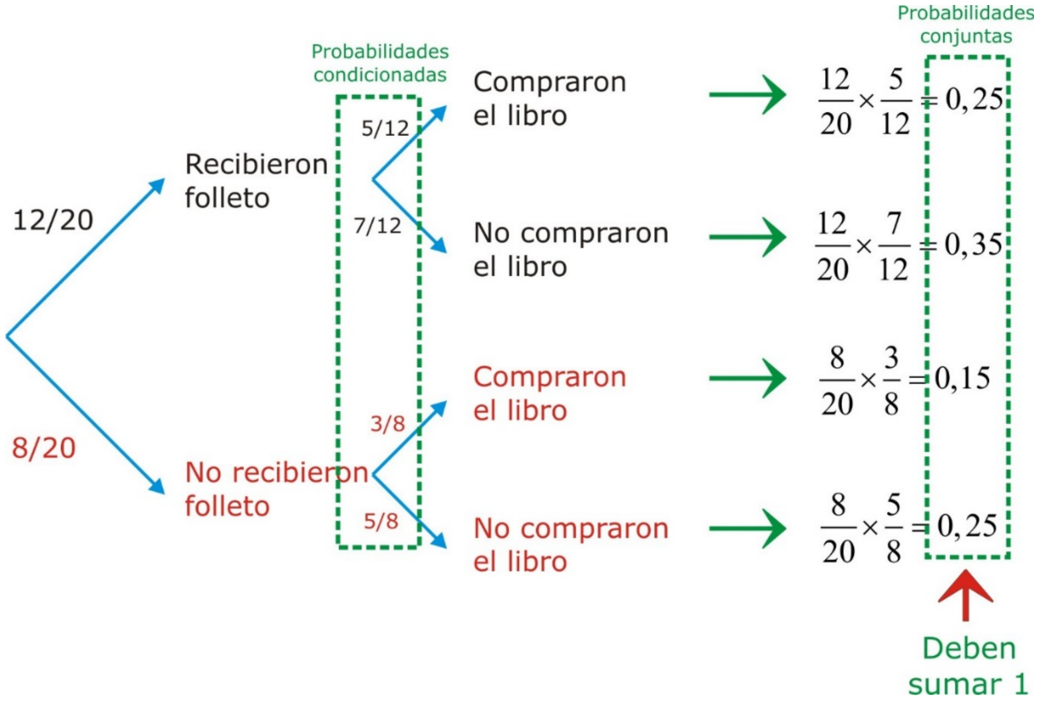
Donde:

- $P(A)$ = Probabilidad del evento A
- $P(\bar{A})$ = Probabilidad del evento No A
- $P(B|A)$ = Probabilidad del evento B condicionado a la ocurrencia previa del evento A
- $P(\bar{B}|A)$ = Probabilidad del evento No B condicionado a la ocurrencia previa del evento A
- $P(B|\bar{A})$ = Probabilidad del evento B condicionado a la ocurrencia previa del evento No A
- $P(\bar{B}|\bar{A})$ = Probabilidad del evento No B condicionado a la ocurrencia previa del evento No A

Ejemplo:

Represente mediante un diagrama de árbol la siguiente situación:

Un autor, por intermedio de la editorial envía folletos promocionando su libro de estadística a 12 de los 20 profesores que enseñan la asignatura en las universidades de la ciudad de Huancayo. Un mes después se constató que 5 de los 12 profesores que recibieron el folleto compraron el libro. En tanto que 3 de los 8 profesores que no recibieron el folleto también compraron el libro.



2. Partición de un espacio muestral

Los eventos A_1, A_2, . . ., A_n forman una partición del espacio muestral Ω, si cumplen las siguientes condiciones:

- Los eventos A_i son mutuamente excluyentes,
- La unión de los eventos A_i es el espacio muestral Ω
- $P(A_i) > 0$

La partición se puede representar gráficamente mediante el diagrama de Venn-Euler en la forma siguiente:

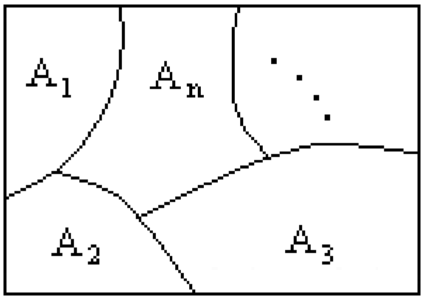


Figura 14. Partición de un espacio muestral
Fuente: Elaboración propia

Ejemplo 1:

Se lanza un dado y sean los eventos: A: El resultado es par y B: El resultado es impar. Comprobar si los eventos A y B son una partición del espacio muestral Ω.

Solución

Tenemos que Ω = {1, 2, 3, 4, 5, 6},

A = {2, 4, 6} y B = {1, 3, 5}.

- Los eventos A y B son mutuamente excluyentes: $A \cap B = \phi$,
- La unión de los eventos A y B es igual al espacio muestral Ω: $A \cup B = \Omega$
- P(A) > 0 y P(B) > 0

Entonces A y B es una partición del espacio muestral.

Ejemplo 2:

En el experimento de lanzar 3 monedas simultáneamente:

Ω = {CCC, CCS, SCC, CSC, CSS, SCS, SSC, SSS}

Se definen los eventos:

A = {CCC}

B = {CCS, SCC, CSC}

C = {CSS, SCS, SSC}

D = {SSS}

¿Forman una partición del espacio muestral?

Solución

- Los eventos A, B, C y D son mutuamente excluyentes: $A \cap B \cap C \cap D = \phi$
- La unión de los eventos A, B, C y D es igual al espacio muestral Ω: $A \cup B \cup C \cup D = \Omega$
- P(A) > 0, P(B) > 0, P(C) > 0, P(D) > 0

Entonces A, B, C y D forman una partición del espacio muestral.

3. Probabilidad total

Sean $A_1, A_2, A_3, \ldots, A_n$ una partición del espacio muestral Ω, entonces para cualquier evento B definido en Ω, se cumple que:

$$P(B) = \sum_{i=1}^{n} P(A_i).P(B \mid A_i)$$

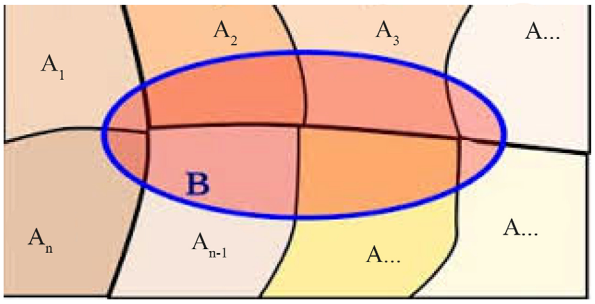
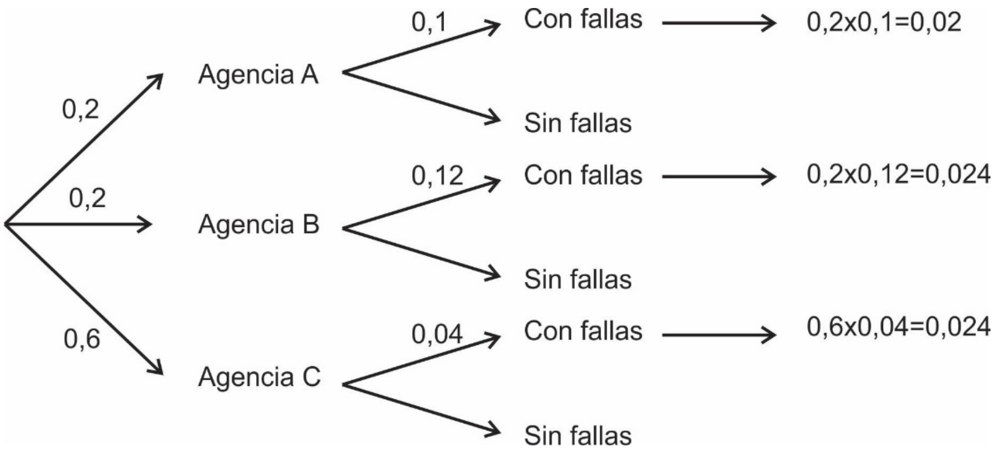


Figura 15. Probabilidad total
Fuente: Elaboración propia

Ejemplo:

Una compañía minera alquila cargadores frontales a tres agencias especializadas: El 20% de la agencia A, el 20% de la agencia B y el 60% de la agencia C. Si 10% de los cargadores de la agencia A, 12% de los provenientes de la agencia B y 4% de los provenientes de la agencia C tienen fallas mecánicas. ¿Cuál es la probabilidad de que se alquile un cargador frontal con fallas mecánicas?



P(Con fallas) = 0,02+0,024+0,024 = 0,068

4. Teorema de Bayes

En el siglo XVIII el reverendo Thomas Bayes, ministro presbiteriano inglés, se planteó la siguiente pregunta: ¿En verdad existe Dios? Estando interesado en las matemáticas intentó desarrollar una fórmula para llegar a evaluar la probabilidad de que Dios existe, con base en la evidencia de la que él disponía aquí en la Tierra. Más adelante, a la muerte de Bayes, otro matemático llamado Laplace mejoró este trabajo y le dio el nombre de Teorema de Bayes.

Simbólicamente, este teorema se presenta de la siguiente manera:

$$P(A_i|B)=\frac{P(A_i)\times P(B|A_i)}{P(A_1).P(B|A_1)+P(A_2).P(B|A_2)+P(A_3).P(B|A_3)+\ldots+P(A_k).P(B|A_k)}$$

Observación:

Si los sucesos A_1, A_2, …., A_k se consideran casos posibles del suceso B, entonces Bayes permite determinar la probabilidad de que uno de los A_i sea causa de la ocurrencia de B.

Ejemplo 1:

Supóngase que en el ejercicio anterior de los cargadores frontales, se ha alquilado un cargador que resultó tener fallas mecánicas. ¿Cuál es la probabilidad de que haya sido alquilado a la agencia C?

Solución:

La probabilidad de que el cargador con fallas mecánicas sea de la compañía C se obtiene aplicando el teorema de Bayes:

$$P(Agencia\ C \mid Con\ \ fallas) = \frac{0{,}6(0{,}04)}{0{,}2(0{,}1)+0{,}2(0{,}12)+0{,}6(0{,}04)} = \frac{0{,}024}{0{,}068} = 0{,}353$$

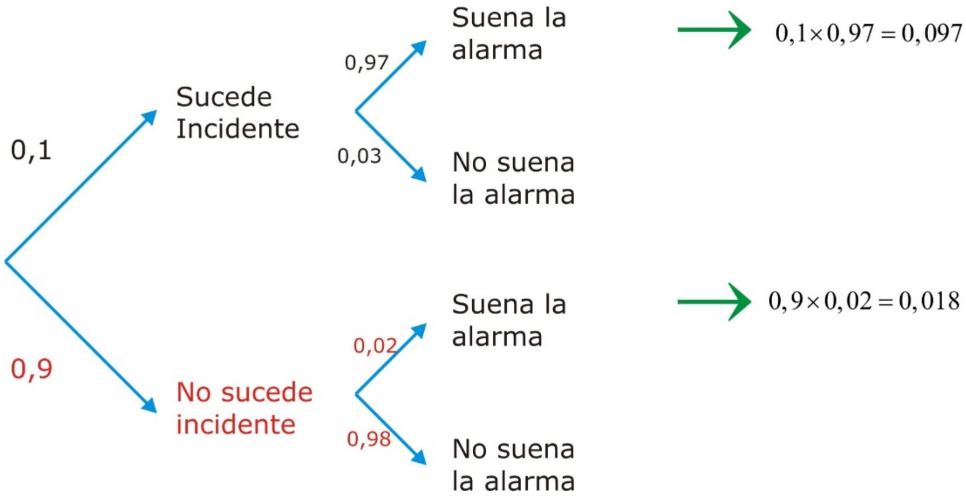
Ejemplo 2:

La probabilidad de que haya un accidente en una fábrica que dispone de alarma es 0,1. La probabilidad de que suene ésta si se ha producido algún incidente es de 0,97 y la probabilidad de que suene si no ha sucedido ningún incidente es 0,02.

En el supuesto de que haya funcionado la alarma, ¿cuál es la probabilidad de que no haya habido ningún incidente?

Solución:

Llevando los datos a un diagrama de árbol:



$$P(Suene) = 0{,}097 + 0{,}018 = 0{,}115 \quad (\text{Prob. } Total)$$

$$P(No \quad incidente \mid Suena \quad alarma) = \frac{0{,}018}{0{,}115} = 0{,}157 \ \ (T.\ de\ Bayes)$$

ACTIVIDAD FORMATIVA Nº 3

Resuelve las actividades de la guía de práctica del texto, haciendo uso del laboratorio de cómputo

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

Probabilidad total y teorema de Bayes:

1. Tres máquinas A, B y C producen respectivamente el 60%, 30% y 10% del número total de artículos de una fábrica. Los porcentajes de desperfectos de producción de estas máquinas son respectivamente 5%; 9% y 11%. Al seleccionarse un artículo resulto bueno. Hallar la probabilidad de que haya sido producido por la máquina C.

2. El 20% de los estudiantes de la UCCI son de postgrado, en tanto que el 80% son de pregrado. La probabilidad de que un estudiante de postgrado sea casado es 0,5 en tanto que para uno de pregrado es sólo 0,1. Se elige al azar un estudiante de la UCCI.

    - ¿Cuál es la probabilidad de que sea casado?
    - Dado que el estudiante es casado, ¿cuál es la probabilidad de que sea de postgrado?

3. La probabilidad de que Alicia estudie para su examen final de Estadística es de 0,2. Si estudia, la probabilidad de que apruebe el examen es de 0,8, en tanto que si no estudia la probabilidad de aprobar es de sólo 0,5.

    - ¿Cuál es la probabilidad de que Alicia apruebe su examen final de Estadística?
    - Dado que Alicia aprobó el examen, ¿cuál es la probabilidad de que ella haya estudiado?

4. Todas las noches el señor Herrera llega tarde a su casa, la señora Herrera que es buena esposa, le deja encendida la luz de la entrada de la casa. La probabilidad de que el señor Herrera llegue borracho es 0,6, si llega borracho, hay una probabilidad de 0,9 de que olvide apagar la luz, en tanto que ésta es sólo de 0,05 si llega sobrio.

    - ¿Cuál es la probabilidad de que el señor Herrera apague la luz en una noche cualquiera?
    - Dado que el señor Herrera apagó la luz una cierta noche, ¿cuál es la probabilidad de que haya llegado borracho?

5. La doctora Castro ha estado enseñando Estadística básica muchos años. Ya sabe que el 80% de sus alumnos hacen toda la tarea. También sabe que el 90% de los que hacen toda la tarea aprueban. De los estudiantes que no hacen la tarea, el 60% aprueba. Luis Laime llevó el curso el semestre pasado con la doctora Castro y obtuvo calificación aprobatoria. ¿Cuál es la probabilidad de que él haya hecho todas sus tareas?

6. En la Cooperativa de Ahorro y Crédito “Tumán” la gerencia sabe por experiencia que la probabilidad de que un socio pague a tiempo su préstamo es de 0,8; además sabe que el 45% de los préstamos pagados a tiempo han sido para financiar compra de artefactos eléctricos y el 65% de los préstamos no pagados a tiempo han sido para el mismo fin. Calcular la probabilidad de que un préstamo que se haya hecho para financiar la compra de un artefacto eléctrico no se pague a tiempo.

7. Basado en su experiencia, el encargado del almacén de una planta de ensamblaje de computadoras ha indicado que en un lote de chips, el 30% pueden venir de la planta A, el 50% puede venir de la planta B y el resto de la planta C. Por otro lado se conoce que: El 1% de los chips que provienen de la planta A son defectuosos, el 2% de los que provienen de la planta B son defectuosos y el 3% de los que provienen de la planta C son defectuosos. Al seleccionar un chip del lote para ser utilizado en un ensamblaje, éste resultó defectuoso. A la luz de este hecho, ¿cuál es la probabilidad de que dicho chip provenga de la planta A?

8. Un laboratorio somete a los choferes que sufren accidentes de tránsito a la prueba del alcoholímetro. ¿Cuál es la probabilidad de que un chofer de esta población esté ebrio, dado que el resultado del alcoholímetro fue positivo?, si se ha determinado que cuando un chofer está ebrio, la prueba proporciona resultado positivo en el 95% de los casos, cuando un chofer no está ebrio, la prueba proporciona resultado negativo en el 94% de los casos, el 2% de los conductores que sufren accidentes manejan ebrios.

9. La facultad de Ingeniería Estadística de cierta universidad, está conformada por tres carreras profesionales: Estadística e Informática, Sistemas y Electrónica. Solamente el 10% de los estudiantes de Estadística e Informática, el 6% de los estudiantes de Sistemas y el 8% de los de Electrónica culminan exitosamente sus estudios. Se sabe que el 50% de los estudiantes son de Estadística e Informática, el 30% de Sistemas y el 20% Electrónica. Si se selecciona al azar un estudiante y éste nos dice que ha culminado exitosamente su carrera, ¿cuál es la probabilidad de que sea de Estadística e Informática?

10. En una investigación se ha detectado que 2 de cada 500 personas adultas padecen de cierta enfermedad, para la cual se ha desarrollado una prueba de diagnóstico. Si a una persona que padece dicha enfermedad se le realiza la prueba de diagnóstico, presentará un resultado positivo el 97% de las veces, mientras que una persona que no padece de dicha enfermedad mostrará un resultado positivo sólo el 2% de las veces. Si se selecciona una persona al azar y es diagnosticada positiva, ¿cuál es la probabilidad de que en verdad no tenga dicha enfermedad?

VIDEOS.

Probabilidad condicional.

https://www.youtube.com/watch?v=Ik9NPdNgXhQ

Teorema de Bayes

https://www.youtube.com/watch?v=vX0c3M1qZIs

LECTURA SELECCIONADA No 1

Historia y relevancia de la teoría de la probabilidad

Estadística para Administración y Economía.

Richard Levin. Pág. 128

Historia y relevancia de la teoría de la probabilidad

Jacob Bernoulli (1654-1705), Abraham de Moivre (1667-1754), el reverendo Thomas Bayes (1702- 1761) y Joseph Lagrange (1736-1813) desarrollaron fórmulas y técnicas para el cálculo de la probabilidad.

En el siglo XIX, Pierre Simon, marqués de Laplace (1749-1827), unificó todas estas ideas y compiló la primera teoría general de probabilidad.

La teoría de la probabilidad fue aplicada con éxito en las mesas de juego y, lo que es más importante en nuestro estudio, a problemas sociales y económicos. La industria de seguros, que surgió en el siglo XIX, requería un conocimiento preciso acerca de los riesgos de pérdida, con el fin de calcular las primas. Medio siglo más tarde, muchos centros de aprendizaje estaban estudiando la probabilidad como una herramienta para el entendimiento de los fenómenos sociales. En la actualidad, la teoría matemática de la probabilidad es la base para las aplicaciones estadísticas, tanto en investigaciones sociales como en la toma de decisiones.

La probabilidad constituye parte importante de nuestra vida cotidiana. En la toma de decisiones personales y administrativas, nos enfrentamos a la incertidumbre y utilizamos la teoría de la probabilidad, admitamos o no el uso de algo tan complejo. Cuando escuchamos una predicción de 70% de posibilidades de lluvia, cambiamos nuestros planes de salir de día de campo y nos quedamos en casa divirtiéndonos con juegos de mesa. Cuando jugamos al *bridge* hacemos algunas estimaciones de

probabilidad antes de intentar una jugada arriesgada. Los administradores que se encargan de inventarios de ropa de moda para mujer deben preguntarse sobre las posibilidades de que las ventas alcancen o excedan un cierto nivel. Antes de la tan publicitada pelea de Muhammed Alí contra Leon Spinks, se afirmaba que Alí había dicho: "Les **apuesto** a que todavía seré el más grande cuando termine la pelea." Y cuando usted mismo empiece a estudiar para el examen del contenido de este libro, seguramente se preguntará: ¿cuál es la posibilidad de que el profesor nos pregunte algo sobre la historia de la teoría de la probabilidad?

Vivimos en un mundo incapaz de predecir el futuro con certeza. Nuestra necesidad de encarar a la incertidumbre nos lleva a estudiar y utilizar la teoría de la probabilidad. En muchos casos, nosotros, como ciudadanos preocupados, tendremos algún conocimiento sobre los posibles resultados de una decisión. Al organizar esta información y considerarla de manera sistemática seremos capaces de reconocer nuestras suposiciones, comunicar nuestro razonamiento a otras personas y tomar una decisión más sólida de aquella que se tomaría si sólo diéramos palos de ciego.

LECTURA SELECCIONADA No 2

¿Debe preocuparse de que le realicen una prueba de detección de drogas cuando solicite un trabajo?

Estadística. Mario Triola. Pág. 137

¿Debe preocuparse de que le realicen una prueba de detección de drogas cuando solicite un trabajo?

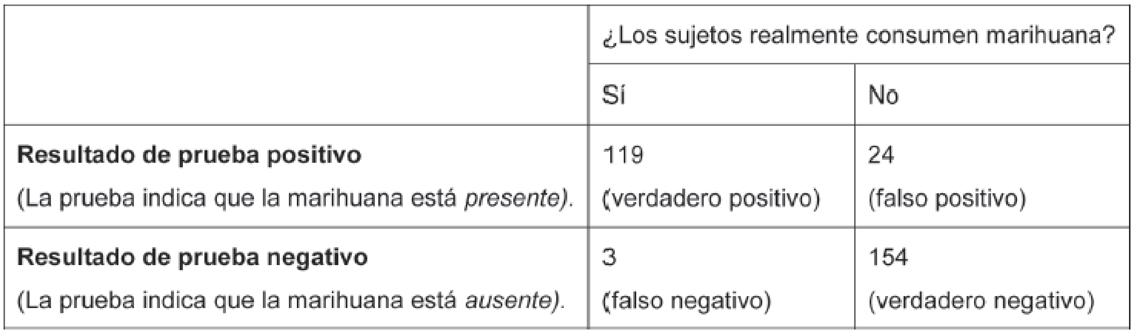
Según la American Management Association, alrededor del 70% de las empresas estadounidenses realizan pruebas de detección de drogas al menos a algunos empleados y aspirantes. El U.S. National Institute on Drug Abuse afirma que aproximadamente el 15% de los jóvenes entre 18 y 25 años consumen drogas ilegales. Quest Diagnostic estima que el 3% de la fuerza laboral general de Estados Unidos consume marihuana. Supongamos que usted solicitó un empleo, tiene excelentes aptitudes (las cuales incluyen la aprobación exitosa de un curso de estadística), le hicieron una prueba de consumo de marihuana y no le dieron el empleo. Usted podría sospechar que no pasó el examen de marihuana, aun cuando no consume esta droga.

Análisis de los resultados:

La tabla muestra los resultados de la prueba "1-Panel- THC" para identificar el consumo de marihuana. Este dispositivo de prueba cuesta $5,95 y la empresa Drug Test Success lo distribuye. Los resultados de la prueba fueron confirmados con cromatografía de gases y espectrometría de masas, que la empresa describe como "el método de confirmación preferido". (Esos resultados se basan en el uso de 50 ng/mL como nivel de corte para determinar la presencia de marihuana). Con base en los resultados de la tabla 4-1, ¿qué probabilidades hay de que la prueba indique que usted consumió marihuana, aunque no sea así? Cuando una prueba muestra la presencia de alguna condición, como una enfermedad o los residuos de alguna droga, se dice que el resultado de la prueba es positivo. Cuando la prueba indica un resultado positivo, pero la condición en realidad no está presente, el resultado es un falso positivo. Es decir, un falso positivo es un error en el que la prueba indica la presencia de una condición, cuando en realidad esta última no se presenta. En este caso, el aspirante al empleo podría sentirse angustiado por la probabilidad de un resultado falso positivo, ya que sería un error que provocaría de manera injusta la negación del empleo. (El contratante podría sentirse preocupado por otro tipo de error, un falso negativo, que se presenta cuando una prueba indica que el aspirante no consume marihuana, cuando en realidad sí lo hace. Este falso negativo podría causar la contratación de un individuo que consume marihuana, y este error puede ser grave para algunos trabajos, como los que realizan los pilotos, los cirujanos o los ingenieros de trenes).

En este capítulo analizaremos preguntas relevantes como éstas: Dados los resultados muestrales de la tabla, ¿cuál es la probabilidad de un resultado falso positivo? ¿Cuál es la probabilidad de un resultado falso negativo? ¿Esas probabilidades son lo suficientemente bajas como para que los aspirantes y los contratantes no se preocupen por tomar decisiones incorrectas motivadas por resultados erróneos de las pruebas?

Tabla
Resultados de exámenes sobre el consumo de marihuana



	¿Los sujetos realmente consumen marihuana?	
	Sí	No
Resultado de prueba positivo (La prueba indica que la marihuana está *presente*).	119 (verdadero positivo)	24 (falso positivo)
Resultado de prueba negativo (La prueba indica que la marihuana está *ausente*).	3 (falso negativo)	154 (verdadero negativo)

GLOSARIO DE LA UNIDAD

E

Evento: Conjunto de uno o más resultados de un experimento. Por ejemplo, un evento consiste en el conjunto de números pares en el lanzamiento de un dado no cargado.

Experimento: Actividad que se observa o se mide. Por ejemplo, un experimento puede consistir en contar el número de respuestas correctas a una pregunta.

I

Independiente: La incidencia de un evento no influye en la probabilidad de que ocurra otro evento.

M

Mutuamente excluyente: La ocurrencia de un evento significa que ninguno de los otros eventos puede ocurrir al mismo tiempo.

P

Probabilidad: Valor entre 0 y 1, inclusive, que indica la posibilidad de que ocurra un evento.

Probabilidad clásica: Probabilidad basada en el supuesto de que cada uno de los resultados tiene la misma probabilidad. De acuerdo con este concepto de probabilidad, si hay n resultados posibles, la probabilidad de un resultado es de 1/n. Por lo tanto, cuando se lanza una moneda al aire, la probabilidad de que salga una cara es de 1/n = 1/2.

Probabilidad condicional: Posibilidad de que un evento ocurra dado que haya ocurrido ya otro evento.

Probabilidad empírica: Concepto probabilístico asentado en la experiencia previa.

Probabilidad subjetiva: Posibilidad de que suceda un evento con base en cualquier información disponible: presentimiento, opinión personal, opiniones de otros, rumores, etcétera.

R

Regla especial de la adición: Para que esta regla sea aplicable, los eventos deben ser mutuamente excluyentes. En el caso de dos eventos, la probabilidad de que ocurran A o B se determina mediante la fórmula P(A o B) = P(A) + P(B)

Regla especial de la multiplicación: Si dos eventos no se encuentran relacionados —son independientes—, se aplica esta regla para determinar la probabilidad de que sucedan al mismo tiempo. P(A y B) = P(A)P(B)

Regla general de la adición: Se utiliza para determinar las probabilidades de eventos complejos compuestos por A o B. P(A o B) = P(A) + P(B)–P(A y B)

Regla general de la multiplicación: Se utiliza para determinar probabilidades de eventos A y B, los cuales se presentan al mismo tiempo. P(A y B) = P(A)P(B|A)

T

Teorema de Bayes: Formulado por el reverendo Bayes en el siglo VIII, está diseñado para determinar la probabilidad de que ocurra un evento A, dado que haya ocurrido otro evento B.

BIBLIOGRAFÍA DE LA UNIDAD III

Devore, J. (2008). *Probabilidad y Estadística* (7 ed.). México: Editorial Cengage Learning Editores S.A.

Triola, M. (2009). *Estadística*. (10 ed.). México: Editorial Pearson.

AUTOEVALUACION N.º 3

1. Elabore el espacio muestral de los siguientes experimentos y calcule la probabilidad de los eventos que se indican:

Experimento: Lanzar un dado: Evento A = Obtener un número mayor o igual que 4	Experimento: Lanzar dos veces una moneda no cargada. Evento: B = Obtener cara y sello (sin importar el orden)

2. Calcular la probabilidad de que una pareja con tres hijos tenga menos de dos varones.

3. De una baraja de 52 cartas se va a extraer una carta. ¿Cuál es la probabilidad de que sea diamante o as?

4. En una urna se tiene 18 bolas rojas, 10 azules y 8 amarillas. ¿Cuál es la probabilidad de que al extraer 3 bolas sin reposición, la primera sea roja, la segunda amarilla y la tercera roja?

5. Tres fábricas: Alfa, Beta y Gamma producen respectivamente el 45%, 25% y 30% del número total de cuchillas termoeléctricas de un país. Los porcentajes de desperfectos de producción de estas fábricas son respectivamente 8%; 5% y 6%. Al seleccionarse al azar una cuchilla termoeléctrica resultó defectuosa. Hallar la probabilidad de que haya sido producido por la fábrica Beta.

PRUEBA DE DESARROLLO – UNIDAD III

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura. De ser necesario vuelve a estudiar los ejemplos desarrollados.

INSTRUCCIONES:

Lea e interprete los diferentes contenidos y ejercicios presentados en los temas anteriores para que resuelva los ejercicios de la presente actividad:

1. Subraye los números que sí representan una probabilidad.

   3/7 ; -0,45 ; 3/4 ; 2,479x10-3 ; 1,576x102 ; 0 ; 1

2. Sea el experimento de lanzar dos dados simultáneamente. Calcule la probabilidad de que los dos lados muestren el mismo número.

3. De 200 niños en extrema pobreza examinados por un médico, se encontró que 80 de ellos padecían de desnutrición, 50 padecían de enfermedades respiratorias y 14 padecían de ambas enfermedades. ¿Cuál es la probabilidad de que al elegir un niño al azar padezca de desnutrición o de enfermedades respiratorias o de ambas enfermedades?

4. Una urna contiene 6 bolas blancas y 8 negras, se extraen sucesivamente y sin reposición 2 bolas. ¿Cuál es la probabilidad de que ambas sean blancas?

5. Una compañía minera alquila cargadores frontales a tres agencias especializadas: El 20% de la agencia A, el 20% de la agencia B y el resto de la agencia C. Si 10% de los cargadores de la agencia A, el 12% de los provenientes de B y 4% de los provenientes de C tienen fallas mecánicas. Se alquila un cargador frontal y resulta que no presenta fallas mecánicas, ¿cuál es la probabilidad que haya sido rentado de la empresa C?

UNIDAD IV

“DISTRIBUCIONES DE PROBABILIDAD”

DIAGRAMA DE PRESENTACIÓN DE LA UNIDAD IV

RESULTADO DE APRENDIZAJE

Al finalizar la unidad, el estudiante será capaz de interpretar las medidas estadísticas de una encuesta en una organización.

UNIDAD IV

TEMA Nº 3

CONTENIDOS	ACTIVIDADES FORMATIVAS (HABILIDADES Y ACTITUDES)	SISTEMA DE EVALUACIÓN (TÉCNICAS Y CRITERIOS)
TEMA Nº 1: DISTRIBUCIONES DE PROBABILIDAD 1 Variable aleatoria 2 Distribución de probabilidad **TEMA Nº 2: DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLE ALEATORIA DISCRETA** 1 Distribuciones de probabilidad para variables aleatorias discretas **TEMA Nº 3: DISTRIBUCIONES DE PROBABILIDAD NORMAL** 1 Distribución normal 2 Distribución normal estándar 3 Puntuaciones "z" 4 Cálculo de áreas bajo la curva normal estándar 5 Aplicaciones de la distribución normal	· Identifica el tipo de variable aleatoria y construye la distribución de la función de probabilidad. · Identifica la distribución de probabilidad discreta de un experimento aleatorio y calcula e interpreta la probabilidad del evento de interés. · Identifica la distribución de probabilidad continua de un experimento aleatorio y calcula e interpreta la probabilidad de distribuciones normales. · Resuelve las actividades de la guía de práctica del texto, haciendo uso del laboratorio de cómputo.	Procedimientos e indicadores de evaluación permanente · Entrega puntual de trabajos realizados. · Calidad, coherencia y pertinencia de contenidos desarrollados. · Prueba teórico-práctica, individual. · Actividades desarrolladas en sesiones tutorizadas Criterios para evaluar habilidades de interpretación, de medidas estadísticas de una encuesta. · Construye e identifica la distribución de la función de probabilidad. · Identifica la distribución de la probabilidad discreta de un experimento aleatorio · Calcula e interpreta la probabilidad del evento de interés. · Identifica la distribución de probabilidad continua de un experimento aleatorio · Calcula e interpreta la probabilidad de distribución.

RECURSOS:

Videos:

Tema Nº 1

Distribuciones de probabilidad

https://www.youtube.com/watch?v=prOG_KfMYtw

Tema Nº 2

Distribución binomial

https://www.youtube.com/watch?v=G1w5gZZaAOg

Distribución Poisson

https://www.youtube.com/watch?v=ID7RPownwBg

Tema Nº 3

Distribución normal: Conceptos y propiedades

https://www.youtube.com/watch?v=b_Oee84PrGg

Lectura complementaria:

Lectura Seleccionada N° 1
¿Qué es una distribución de probabilidad?
Estadística para Administración y Economía.
Richard Levin. Pág. 128

Lectura Seleccionada N° 2
¿Yo mido 1,68. ¿Por qué la probabilidad de que una estatura sea 1,68 calculada con la distribución Normal es 0?
55 Respuestas a dudas estadísticas.
Roberto Behar Gutiérrez, Pere Grima Cintas. Página 51

✓ INSTRUMENTO DE EVALUACIÓN	Prueba de desarrollo
BIBLIOGRAFÍA (BÁSICA Y COMPLEMENTARIA)	**BÁSICA** Devore, J. (2008). *Probabilidad y Estadística* (7 ed.). México: Editorial Cengage Learning Editores S.A. **COMPLEMENTARIA** Ross, S. (2001). *Probabilidad y Estadística para Ingenieros* (3 ed.). México: Mc Graw Hill. Berenson, M. & Levine, D. (2010). *Estadística Básica en Administración, Conceptos y aplicaciones*. México: Prentice Hall.
RECURSOS EDUCATIVOS DIGITALES	· Instituto Nacional de Estadística e Informática. Disponible en: http://www.inei.gob.pe/ · Tokio explora en un congreso la "imprescindible" ciencia de la estadística. (2012, Jul 04). EFE News Service. Disponible en: http://search.proquest.com/docview/1023206430?accountid=146219 · Canales, E. (2005, Jul 26). Mexicar / AMLO sin estadistica. El Norte. Disponible en:: http://search.proquest.com/docview/311786309?accountid=146219

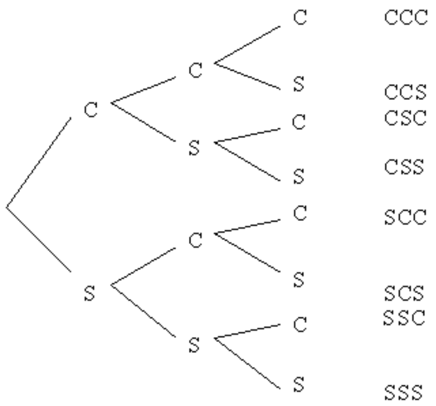
TEMA Nº 1:
DISTRIBUCIONES DE PROBABILIDAD

Estimado estudiante, hemos llegado a la última unidad de la asignatura. Para este tema debes tener en cuenta que las distribuciones de probabilidad están relacionadas con la distribución de frecuencias. De hecho, podemos pensar en la distribución de probabilidad como una distribución de frecuencias teórica. Una distribución de frecuencias teórica es una distribución de probabilidades que describe la forma en que se espera que varíen los resultados. Debido a que estas distribuciones tratan sobre expectativas de que algo suceda, resultan ser modelos útiles para hacer inferencias y tomar decisiones de incertidumbre.

1. Variable aleatoria:

Webster (2001) afirma que, una variable aleatoria es una variable cuyo valor es el resultado de un evento aleatorio.

Por ejemplo, supongamos que se lanza una moneda tres veces y se anota el número de caras que se obtienen. Los posibles resultados se muestran a continuación:



Ω = {ccc, ccs, csc, css, scc, scs, ssc, sss}

Se puede observar que los posibles resultados son 0 caras (sss), 1 cara (css, scs, ssc), 2 caras (ccs, csc, scc) o 3 caras (ccc). La variable aleatoria es el número de caras que puede obtenerse y los posibles resultados son los valores de la variable aleatoria.

Como segundo ejemplo, el contenido de agua mineral en botellas oscilaba aleatoriamente entre 618 a 625 mililitros.

Los contenidos reales de las botellas, en mililitros, son los valores de la variable aleatoria.

Como sugieren los ejemplos, las variables aleatorias pueden ser discretas o continuas:

- **Variable aleatoria discreta**: Es aquella que puede asumir sólo ciertos valores, con frecuencia números enteros que resultan principalmente del conteo. Estos valores constituyen un conjunto finito o bien pueden ser puestos en lista en una secuencia infinita asociada a un proceso de conteo. Por ejemplo, el número de caras en el lanzamiento de tres monedas, el número de huevos que puede poner una gallina, el número de piezas defectuosas que se produzcan en cierta industria, etc.
- **Variable aleatoria continua**: Es aquella que puede tomar cualquier valor dentro de un rango dado de infinitos valores. Provienen generalmente de la medición. Por ejemplo el contenido de agua mineral de una botella, la cantidad de leche que produce una vaca, el peso de un niño al nacer, etc.

2. Distribución de probabilidad:

Para Webster (2001) es un despliegue de todos los posibles resultados de un experimento junto con las probabilidades de cada resultado. Así mismo, Triola (2009) afirma que una distribución de probabilidad puede presentarse como tabla, gráfico o fórmula.

La probabilidad de que la variable aleatoria X tome algún valor específico x_i se escribe $P(X=x_i)$.

2.1. Requisitos de una distribución de probabilidad:

- $\sum P(x)=1$ donde "x" asume todos los posibles valores. (La suma de todas las probabilidades debe ser 1, pero valores como 0,999 o 1,001 son aceptables porque son el resultado de errores de redondeo).
- $0 \leq P(x) \leq 1$ para cada valor individual de x. (Es decir, cada valor de probabilidad debe ubicarse entre 0 y 1, inclusive).

Ejemplo 1:

En el caso del lanzamiento de tres monedas sobre una mesa se obtuvo que:

Espacio muestral: Ω = {ccc, ccs, csc, css, scc, scs, ssc, sss}

Variable aleatoria X = Número de caras = {0; 1; 2 y 3}

- P(X=0) = 1/8 = 0,125, porque sólo existe un caso donde no aparece ninguna cara (sss) de un total de 8 posibles resultados en el espacio muestral.
- P(X=1) = 3/8 = 0,375, porque existen tres casos donde aparece una cara (css, scs, ssc) de un total de 8 posibles resultados en el espacio muestral.
- P(X=2) = 3/8 = 0,375, porque existen tres casos donde aparecen dos caras (ccs, csc, scc) de un total de 8 posibles resultados en el espacio muestral.
- P(X=3) = 1/8 = 0,125, porque sólo existe un caso donde aparecen tres caras (ccc) de un total de 8 posibles resultados en el espacio muestral.

Luego estos resultados se pueden expresar como una tabla de la siguiente forma:

X=Nº CARAS	P(X=XI)
0	1/8 = 0,125
1	3/8 = 0,375
2	3/8 = 0,375
3	1/8 = 0,125

También se pueden expresar como un gráfico:

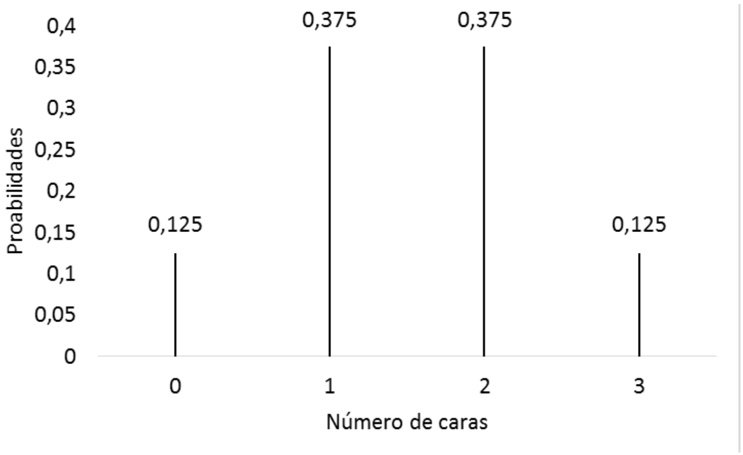


Figura 16. Distribución de probabilidad del lanzamiento de tres monedas
Fuente: Elaboración propia

Comprobando los requisitos de una distribución de probabilidad:

- $\sum P(x)=1$: 0,125 + 0,375 + 0,375 + 0,125 = 1
- $0 \leq P(x) \leq 1$: Todos los valores de probabilidad (0,125; 0,375; 0,375 y 0,125) son números comprendidos entre 0 y 1.

2.2. Media y varianza de las distribuciones de probabilidad:

Así como en la Unidad II se calculó la media, varianza y desviación estándar de un conjunto de datos, también se puede hacer los mismo con una distribución de probabilidad.

- **Media o valor esperado de una distribución de probabilidad**:

$$\mu = E(X) = \sum\left[x.P(x)\right]$$

- **Varianza:**

$$\sigma^2 = \sum\left[(x-\mu)^2 .P(x)\right]$$

- **Desviación estándar:**

$$\sigma = \sqrt{\sigma^2}$$

Obsérvese que la notación de la media (μ), varianza (σ^2) y desviación estándar (σ) de la distribución de probabilidad se expresan como parámetros poblacionales, es decir, con letras griegas minúsculas.

Ejemplo 2:

Calcule la media, varianza y desviación estándar de la distribución de probabilidad dada por el número de caras del lanzamiento de tres monedas.

X=Nº CARAS	P(X=XI)
0 1 2 3	1/8 = 0,125 3/8 = 0,375 3/8 = 0,375 1/8 = 0,125

Solución

X=Nº CARAS	P(X=XI)	$x.P(x)$	$(x-\mu)^2.P(x)$
0 1 2 3	1/8 = 0,125 3/8 = 0,375 3/8 = 0,375 1/8 = 0,125	0 0,375 0,75 0,375	(0–1,5)2.(0,125) = 0,28125 (1–1,5)2.(0,375) = 0,09375 (2–1,5)2.(0,375) = 0,09375 (3–1,5)2.(0,125) = 0,28125
S	1	1,5	0,75

- Media o esperanza: μ =E(x)=1,5 caras
- Varianza: σ^2 =0,75 caras^2
- Desviación estándar: σ =0,866 caras

2.3. Esperanza matemática o valor esperado:

La media es un valor representativo que sirve para representar una distribución de probabilidad. También es el valor promedio a largo plazo de la variable aleatoria. Los nombres de esperanza matemática y valor esperado tienen su origen en los juegos de azar y hacen referencia a la ganancia promedio esperada por un jugador cuando hace un gran número de apuestas.

Si la esperanza matemática es cero, E(x) = 0, el juego es equitativo, es decir, no existe ventaja ni para el jugador ni para la casa de juego.

ACTIVIDAD FORMATIVA Nº 1

Identifica el tipo de variable aleatoria y construye la distribución de la función de probabilidad.

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

1. Identifique si las siguientes variables aleatorias son discretas o continuas:
   - La estatura de un jugador de basquetbol de la NBA, que se selecciona aleatoriamente.
   - El número de puntos que anota un jugador de basquetbol de la NBA, seleccionado aleatoriamente.
   - El tiempo exacto de juego de un jugador de basquetbol de la NBA, seleccionado aleatoriamente.
   - El número de atletas que participaron en la Maratón de Los Andes del año 2011.
   - El salario de un administrador de cualquier institución del sistema bancario del Perú.
   - El costo de la realización de una película que se selecciona aleatoriamente.
   - El número de películas que actualmente se exhiben en los cines del Perú.
   - La duración exacta de una película seleccionada aleatoriamente.
   - El número de actores que aparecen en una película que se selecciona aleatoriamente.
   - El peso del actor principal de una película seleccionada aleatoriamente.

2. Identificación de distribuciones de probabilidad: En los ejercicios que se presentan, determine si se trata de una distribución de probabilidad. En los casos en que no se describa una distribución de probabilidad, identifique los requisitos que no se satisfacen. En los casos en que se describa una distribución de probabilidad, calcule su media y desviación estándar.
   - En un estudio de un método de selección del género en niños, las parejas de un grupo control no reciben tratamiento y cada una de ellas tiene tres hijos. La distribución de probabilidad del número de niñas se presenta en la tabla anexa.

X	P(X)
0	0,125
1	0,375
2	0,375
3	0,125

- Durante la fabricación de DVDs Sony, se seleccionan aleatoriamente grupos de DVD y se calcula el número de defectos x en cada grupo.

X	P(X)
0	0,502
1	0,365
2	0,098
3	0,011
4	0,001

- La tabla adjunta se construye de datos que se obtienen en un estudio del número de videocintas rentadas en Blockbuster.

X	P(X)
0	0,04
1	0,26
2	0,36
3	0,20
4	0,08

- Se selecciona aleatoriamente a un preso, convicto por manejar intoxicado, la distribución de probabilidad del número x de sentencias previas por este delito se describe en la tabla adjunta:

X	P(X)
0	0,512
1	0,301
2	0,132
3	0,055

- La Línea Air America tiene, por rutina, la política de vender un número de boletos que rebasa el número de asientos por cada vuelo, ya que la experiencia demuestra que algunos pasajeros no se presentan. La variable aleatoria x representa el número de pasajeros que no pueden abordar porque hay más pasajeros que asientos.

X	P(X)
0	0,805
1	0,113
2	0,057
3	0,009
4	0,002

3. Considere el experimento que consiste en lanzar una moneda dos veces.
    - Enumere los resultados experimentales.
    - Defina una variable aleatoria que represente el número de caras en los dos lanzamientos.
    - Dé el valor que la variable aleatoria tomará en cada uno de los resultados experimentales.
    - ¿Es una variable aleatoria discreta o continua?

4. Las tres tablas que se presentan a continuación muestran los valores de la "variable aleatoria" y sus "probabilidades". Sin embargo, sólo uno de éstos es una distribución de probabilidad.

a) ¿Cuál es?

X	P(X)
5	0,3
10	0,3
15	0,2
20	0,4

X	P(X)
5	0,1
10	0,3
15	0,2
20	0,4

X	P(X)
5	0,5
10	0,3
15	-0,2
20	0,4

b) Utilizando la distribución de probabilidad correcta, encontrar la probabilidad de que x sea:

- Exactamente 15
- No más de 10
- Más de 5

c) Calcule la media, varianza y desviación estándar de esta distribución.

5. Calcule la media y la varianza de la siguiente distribución de probabilidad discreta:

X	P(X)
0	0,2
1	0,4
2	0,3
3	0,1

6. Calcule la media y la varianza de la siguiente distribución de probabilidad discreta:

X	P(X)
5	0,3
10	0,3
15	0,2
20	0,4

VIDEO:

Distribuciones de probabilidad

https://www.youtube.com/watch?v=qnkoCZhnEwk

TEMA Nº 2:
DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLE ALEATORIA DISCRETA

1. Distribuciones de probabilidad para variables aleatorias discretas:

1.1. Distribución Binomial

Es un tipo de distribución de probabilidad para variables aleatorias discretas que se ocupa de experimentos que según Devore (2008) cumplen las siguientes condiciones:

- El experimento consta de una secuencia de "n" experimentos más pequeños llamados "ensayos", donde "n" se fija antes del experimento.
- Cada ensayo puede dar por resultado sólo uno de dos resultados posibles, los cuales se denotan como éxito (E) y fracaso (F).
- Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.
- La probabilidad de éxito (p) y fracaso (q) son constantes de un ensayo a otro. La probabilidad del fracaso se obtiene como 1-p.

En este tipo de experimento la variable aleatoria X está dada por el número de éxitos que se esperan obtener en los "n" ensayos.

La distribución Binomial puede describirse mediante la siguiente fórmula:

$$P\left(X = x\right) = C_x^n . p^x . q^{n-x}$$

Donde:

$$C_x^n = \frac{n!}{x!\left(n - x!\right)}$$

n = Número de ensayos

p = Probabilidad del éxito

q = Probabilidad del fracaso (q = 1-p)

x = Número de éxitos que se espera obtener en los "n" ensayos

Ejemplo 1:

Se lanza tres veces una moneda al aire. Calcule la probabilidad de obtener 2 caras.

Solución:

Comprobamos si se cumplen las condiciones de un experimento binomial:

El experimento consta de una secuencia de "n" experimentos más pequeños llamados "ensayos", donde "n" se fija antes del experimento.	Se lanza tres veces una moneda al aire. n = 3
Cada ensayo puede dar por resultado sólo uno de dos resultados posibles, los cuales se denotan como éxito (E) y fracaso (F).	E: Obtener cara F: Obtener sello
Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.	Lo que ocurra en un lanzamiento cualquiera de la moneda en nada afecta a lo que pueda ocurrir en los otros lanzamientos.
La probabilidad de éxito (p) y fracaso (q) son constantes de un ensayo a otro. La probabilidad del fracaso se obtiene como 1-p	p = Probabilidad de obtener cara al lanzar una vez la moneda p = 0,5 q = Probabilidad de obtener sello al lanzar una vez la moneda q = 0,5 Estos valores se mantienen constantes en cada ensayo.
La variable aleatoria X en este caso representa el número de éxitos que se esperan en los "n" ensayos	x = 2 Nota: La variable pudo tomar cualquiera de estos valores: 0; 1; 2 y 3.

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X=x)=C_x^n.p^x.q^{n-x}$$

$$P(x=2)=C_2^3.(0,5)^2.(0,5)^1$$

$$P(x=2)=0,375$$

Como se puede apreciar, el resultado es el mismo que se obtuvo mediante la tabla de distribución de probabilidad del ejemplo 2.

Ejemplo 2:

El experimento consiste en extraer una muestra aleatoria de tamaño 4, una a una, con reposición, de un lote de 100 artículos que contiene 5 artículos defectuosos. Calcule la probabilidad de extraer exactamente 1 artículo defectuoso.

Solución:

Comprobamos si se cumplen las condiciones de un experimento binomial:

El experimento consta de una secuencia de "n" experimentos más pequeños llamados "ensayos", donde "n" se fija antes del experimento.	Extraer 4 elementos en la muestra. n = 4
Cada ensayo puede dar por resultado sólo uno de dos resultados posibles, los cuales se denotan como éxito (E) y fracaso (F).	E: Artículo defectuoso F: Artículo no defectuoso

Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.	Como menciona que las extracciones se hacen con reposición, entonces se deduce que los ensayos son independientes.
La probabilidad de éxito (p) y fracaso (q) son constantes de un ensayo a otro. La probabilidad del fracaso se obtiene como 1-p	p = Probabilidad de obtener un artículo defectuoso p = 5/100 = 0,05 q = Probabilidad de obtener un artículo no defectuoso q = 1-0,05 = 0,95 Estos valores se mantienen constantes en cada ensayo.
La variable aleatoria X en este caso representa el número de éxitos que se esperan en los "n" ensayos	x = 1 Nota: La variable pudo tomar cualquiera de estos valores: 0; 1; 2; 3 y 4

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X=x)=C_x^n.p^x.q^{n-x}$$

$$P(x=1)=C_1^4.(0,05)^1.(0,95)^3$$

$$P(x=1)=0,171$$

Ejemplo 3:

El experimento consiste en extraer una muestra aleatoria de tamaño 4, una a una, con reposición, de un lote de 100 artículos que contiene 5 artículos defectuosos. Calcule la probabilidad de extraer *A LO MÁS* 2 artículos defectuosos.

Solución:

Comprobamos si se cumplen las condiciones de un experimento binomial:

El experimento consta de una secuencia de "n" experimentos más pequeños llamados "ensayos", donde "n" se fija antes del experimento.	Extraer 4 elementos en la muestra. n = 4
Cada ensayo puede dar por resultado sólo uno de dos resultados posibles, los cuales se denotan como éxito (E) y fracaso (F).	E: Artículo defectuoso F: Artículo no defectuoso
Los ensayos son independientes, de modo que el resultado en cualquier ensayo particular no influye en el resultado de cualquier otro ensayo.	Como menciona que las extracciones se hacen con reposición, entonces se deduce que los ensayos son independientes.

La probabilidad de éxito (p) y fracaso (q) son constantes de un ensayo a otro. La probabilidad del fracaso se obtiene como 1-p	p = Probabilidad de obtener un artículo defectuoso p = 5/100 = 0,05 q = Probabilidad de obtener un artículo no defectuoso q = 1-0,05 = 0,95 Estos valores se mantienen constantes en cada ensayo.
La variable aleatoria X en este caso representa el número de éxitos que se esperan en los "n" ensayos	En el enunciado dice "a lo más 2 artículos defectuosos" x = 0; x = 1; x = 2 Nota: Es una probabilidad acumulada, es decir, hay que sumar la probabilidad de que "x" tome los valores 0; 1 y 2.

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X=x)=C_x^n.p^x.q^{n-x}$$

$$P(x=0)=C_0^4.(0,05)^0.(0,95)^4=0,81450625$$

$$P(x=1)=C_1^4.(0,05)^1.(0,95)^3=0,171475$$

$$P(x=2)=C_2^4.(0,05)^2.(0,95)^2=0,0135375$$

Para dar la respuesta final sumamos todas las probabilidades: 0,9995

1.2. Distribución Poisson

Según Webster (2000) es una distribución de probabilidad para variables aleatorias de gran utilidad en la medición de la frecuencia relativa de un evento sobre alguna unidad de tiempo o espacio. Con frecuencia se le utiliza para describir el número de llegadas de clientes por hora, el número de accidentes industriales cada mes, el número de conexiones eléctricas defectuosas por kilómetro en un sistema eléctrico de una ciudad, etc.

Asimismo sostiene que son necesarios dos supuestos para la aplicación de la distribución de Poisson:

- La probabilidad de ocurrencia del evento es constante para dos intervalos cualesquiera de tiempo o espacio.
- La ocurrencia del evento en un intervalo es independiente de la ocurrencia en otro intervalo cualquiera.

Devore (2008) manifiesta que no existe un experimento simple en el cual esté basada la distribución de Poisson.

En este tipo de experimento la variable aleatoria X es el número de ocurrencias en un intervalo dado.

La distribución Poisson puede describirse mediante la siguiente fórmula:

$$P(X=x)=\frac{\mu^x.e^{-\mu}}{x!}$$

Donde:

μ = Es la media aritmética del número de ocurrencias en un

intervalo dado

e = Constante matemática épsilon (e = 2,71828)

x = Número de veces que ocurre el evento.

Ejemplo 7:

Un profesor de Estadística anima a sus estudiantes a actuar en forma prudente consultando al tutor si tienen alguna pregunta mientras se preparan para el examen final. Parece que la llegada de los estudiantes a la oficina del tutor se ajusta a una distribución de Poisson, con un promedio de 5,2 estudiantes cada 20 minutos. El profesor está preocupado porque si muchos estudiantes necesitan los servicios del tutor, puede generarse un problema de congestión. ¿Cuál es la probabilidad de que cuatro estudiantes lleguen durante el primer intervalo de 20 minutos?

Solución:

Se sabe que:

μ = 5,2 estudiantes cada 20 minutos

x = 4 estudiantes en 20 minutos

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X = x) = \frac{\mu^x . e^{-\mu}}{x!}$$

$$P(x = 4) = \frac{(5,2)^4 (2,71828)^{-5,2}}{4!}$$

$$P(x = 4) = 0,1681$$

Ejemplo 8:

El número de personas que llegan para atención a una sala de emergencias puede ser modelado mediante un proceso de Poisson, con parámetro de razón de cinco por hora. ¿Cuál es la probabilidad de que ocurran tres arribos de pacientes durante una hora en particular?

Solución:

Se sabe que:

μ = 5 pacientes por hora

x = 3 pacientes en una hora

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X=x)=\frac{\mu^{x}.e^{-\mu}}{x!}$$

$$P(x=3)=\frac{(5)^{3}(2,71828)^{-5}}{3!}$$

$$P(x=4)=0,1404$$

Ejemplo 9:

El número de personas que llegan para atención a una sala de emergencias puede ser modelado mediante un proceso de Poisson, con parámetro de razón de cinco por hora. ¿Cuál es la probabilidad de que ocurran más de dos arribos de pacientes durante una hora en particular?

Solución:

Se sabe que:

μ = 5 pacientes por hora

x = 0; x = 1; x = 2 pacientes en una hora

Calculamos la probabilidad pedida aplicando la fórmula:

$$P(X=x)=\frac{\mu^{x}.e^{-\mu}}{x!}$$

$$P(x=0)=\frac{(5)^{0}(2,71828)^{-5}}{0!}=0,0067379$$

$$P(x=1)=\frac{(5)^{1}(2,71828)^{-5}}{1!}=0,0336897$$

$$P(x=2)=\frac{(5)^{2}(2,71828)^{-5}}{2!}=0,0842243$$

Para dar la respuesta final sumamos dichas probabilidades y calculamos su complementario (porque nos piden la probabilidad de más de dos)

1 – (0,0067379+0,0336897+0,0842243) = 0,8753481

ACTIVIDAD FORMATIVA Nº 2

Identifica la distribución de probabilidad discreta de un experimento aleatorio y calcula e interpreta la probabilidad del evento de interés.

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

DISTRIBUCIÓN BINOMIAL:

1. Dentro de la población de una prisión, el 30% de los reclusos se encuentran allí por delitos menores. Se selecciona una muestra aleatoria de 15. ¿Cuál es la probabilidad de que el número de reclusos por delitos menores en esa muestra sea 4?

2. Aproximadamente 3/5 de las personas de la Urbanización Gonzales pertenecen al grupo sanguíneo ORH. ¿Cuál es la probabilidad de que en una muestra aleatoria de 5 personas: a) 2 pertenezcan a dicho grupo, b) Menos de 4 pertenezcan a dicho grupo.

3. Un examen tipo IBM contenía 20 preguntas y cada una de ellas con 5 respuestas alternativas. Si un alumno desconoce todas las respuestas correctas y contestó su examen al azar. ¿Cuál es la probabilidad de que conteste: a) Correctamente 8 preguntas, b) Todo incorrectamente, c) Correctamente de 8 a 10 preguntas?

4. El 80% de los habitantes que viven en un pueblo joven se dedicaban al comercio ambulatorio. Se seleccionan 7 pobladores al azar. ¿Qué probabilidad hay de que: a) 3 de ellos se dediquen al comercio ambulatorio, b) Todos se dediquen al comercio ambulatorio, c) Ninguno se dedique al comercio ambulatorio?

5. Si el 30% de los recién nacidos en la Maternidad de Huancayo nacen con bajo peso. ¿Cuál es la probabilidad de que de 10 nacimientos que tienen lugar cierto día, 2 tengan bajo peso?

6. Se sabe que en el distrito del Tambo el 75% de los establecimientos comerciales pequeños no entregaban factura al momento de efectuar una transacción económica. ¿Cuál es la probabilidad de que, en una muestra de 16 negocios, a lo más la mitad entregue factura?

7. En base a datos confiables, el 25% de los reclusos de un cierto penal son sentenciados sin prueba alguna. Si se selecciona una muestra de 15 reclusos, ¿cuál es la probabilidad de que no más de 3 reclusos hayan sido sentenciados injustamente?

DISTRIBUCIÓN DE POISSON:

1. La central telefónica de una empresa recibe un promedio de 3,5 órdenes de pedido por hora. Estas ocurrencias se producen al azar. ¿Cuál es la probabilidad de que se produzcan exactamente 4 llamadas en una hora dada?

2. En un experimento de laboratorio, el número promedio de partículas radioactivas que pasan a través de un contador durante un milisegundo es 4. Hallar la probabilidad de que ingresen 6 partículas al contador en un milisegundo determinado.

3. En una investigación se ha determinado que el número de muertes debido a una epidemia ha sido en promedio 4 por día. Suponiendo que el número de muertes sigue una distribución de Poisson, hallar la probabilidad de que en un determinado día mueran 3 personas.

4. Una oficina de ventas recibe un promedio de 5,5 pedidos por hora. Suponiendo que los pedidos siguen

una distribución de Poisson, ¿cuál es la probabilidad de que se realicen menos de 4 pedidos en una hora determinada?

5. El número promedio de litros de combustible que se consumen en un barrio de la localidad es de 8 por día. Hallar la probabilidad de que en un día cualquiera se logre consumir de 6 a 7 litros de combustible.
6. Hasta el momento han llegado camiones a un muelle de carga y descarga en forma aleatoria a una tasa de uno por hora. ¿Cuál es la probabilidad de que lleguen 2 camiones en la próxima hora?
7. La Financiera de Crédito "Edificar" recibe en promedio 2,2 solicitudes de préstamos para mejoramiento de vivienda por semana. ¿Cuál es la probabilidad de que reciba 2 solicitudes esta semana?

VIDEOS

Distribución binomial

www.youtube.com/watch?v=EisaSQ1j_Kk

Distribución Poisson

https://www.youtube.com/watch?v=D-MV0DpDQCY

TEMA Nº 3:
DISTRIBUCIÓN DE PROBABILIDAD NORMAL

1. DISTRIBUCIÓN NORMAL

Es la más importante distribución en toda la probabilidad y estadística. Muchas poblaciones numéricas tienen distribuciones que pueden ser modeladas muy bien por una curva normal. Los ejemplos incluyen estaturas, pesos, características físicas, errores de medición en experimentos científicos, tiempos de reacción de máquinas, mediciones de inteligencia, calificaciones de exámenes, etc.

Para Triola (2009) si una variable aleatoria continua tiene una distribución con una gráfica simétrica y en forma de campana, como la de la figura, y puede expresarse por medio de la fórmula que se presenta, decimos que tiene una distribución normal.

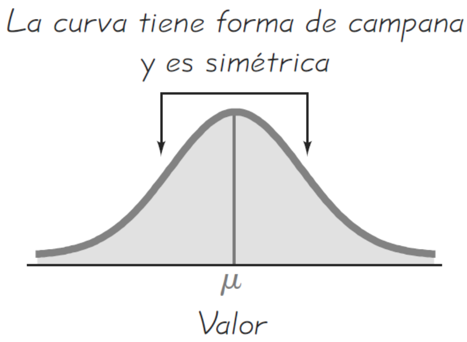


Figura 17. Distribución normal
Fuente: Mario Triola

Cuya fórmula es:
$$y = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

Las letras π y e representan los valores constantes de 3.14159… y 2.71828…, respectivamente.

Los símbolos μ y s representan valores para la media y la desviación estándar, respectivamente. Estos valores vienen a ser los parámetros de la distribución normal y pueden asumir cualquier valor. Por lo tanto, se dice que existe una familia de curvas normales, es decir, infinitas curvas normales dependiendo del valor de sus parámetros.

2. Distribución normal estándar

Triola (2009) menciona que la distribución normal estándar es una distribución normal de probabilidad con μ = 0 y s = 1, y el área total debajo de su curva es igual a 1.

Adicionalmente podemos decir que la curva normal estándar es asintótica al eje “x” y a la vez es simétrica y con forma de campana, por ello se le conoce como la campana de Gauss.

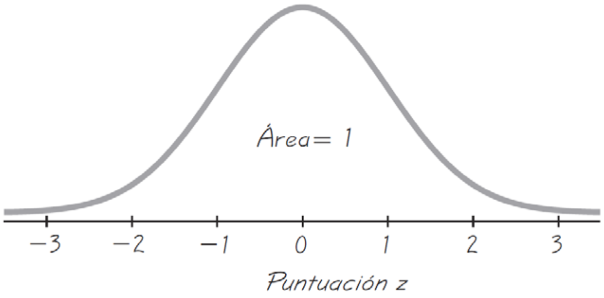


Figura 18: Distribución normal estándar
Fuente: Mario Triola

3. Puntuaciones “z”

Según Triola (2009) una puntuación “z” (o valor estandarizado) es el número de desviaciones estándar que un valor “x” se encuentra por arriba o por debajo de la media. Se calcula utilizando la siguiente expresión:

$$z = \frac{x - Media}{Desv.Estándar}$$

Nota:

- Redondear “z” a dos cifras decimales.
- Siempre que un valor “x” sea menor que la media, su puntuación “z” correspondiente será negativa.
- Siempre que un valor “x” sea mayor que la media, su puntuación “z” correspondiente será positiva.
- Siempre que un valor “x” sea igual que la media, su puntuación “z” correspondiente será cero (0).

Ejemplo 1:

Un grupo de pacientes del consultorio de Cirugía General del Hospital “Carrión” de Cerro de Pasco, tienen un peso promedio de 65 Kg y una desviación estándar de 2,58 kg. Se elige al azar a dos de los pacientes cuyos pesos son de 60 kg y 67 kg. Calcule la puntuación “z” correspondiente al peso de cada paciente e interprete.

Solución

Llamaremos x_1 = 60 kg y x_2 = 67 kg

Calculamos sus puntuaciones “z” mediante la fórmula:

$$z_1 = \frac{60-65}{2,58} = -1,94$$

$$z_2 = \frac{67-65}{2,58} = 0,78$$

Interpretando podemos decir que el paciente cuyo peso es de 60 kg se encuentra a una distancia de 1,94 desviaciones estándar a la izquierda del peso promedio.

En tanto, el paciente cuyo peso es de 60 kg se encuentra a una distancia de 0,78 desviaciones estándar a la derecha del peso promedio.

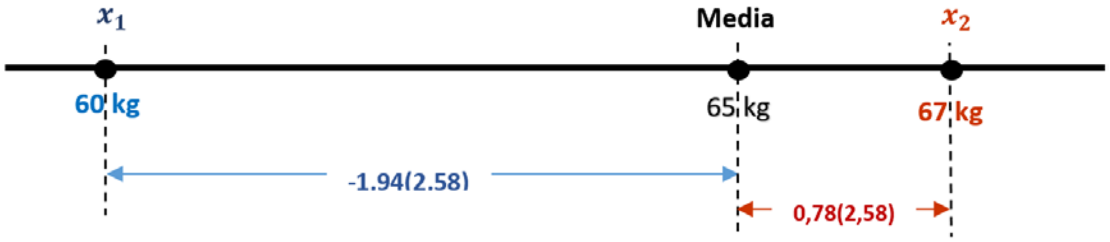
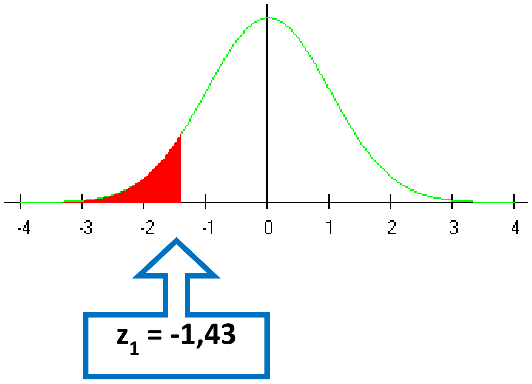


Figura 19. Puntuación "z"
Fuente: Elaboración propia

4. Cálculo de áreas bajo la curva normal estándar:

Primer caso: Área a la izquierda de "z" (se calcula directo de la tabla Anexo "A")

Calcule el área bajo la curva normal a la izquierda de z_1 = -1,43

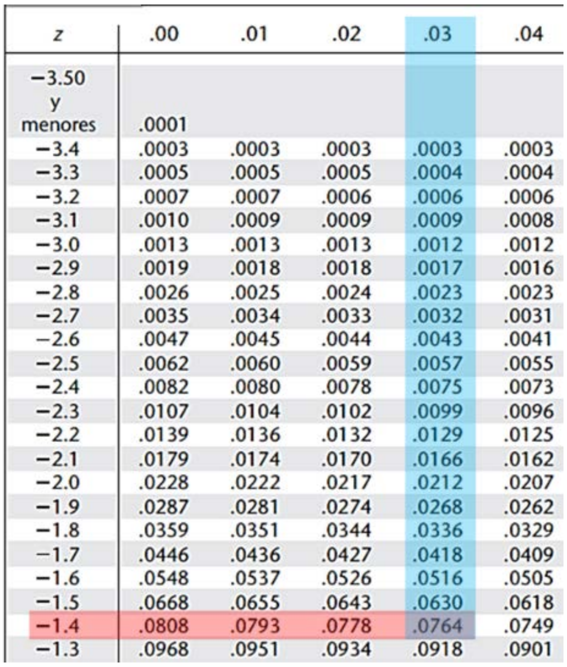


Para calcular dicha área se busca en la tabla "Z NEGATIVA", ya que el valor de "z" es negativo. Luego se busca el valor -1,4 en la columna de "z" y el valor 3 en la fila superior como .03

El área buscada se encuentra en la intersección de la fila con la columna.

Por lo tanto:

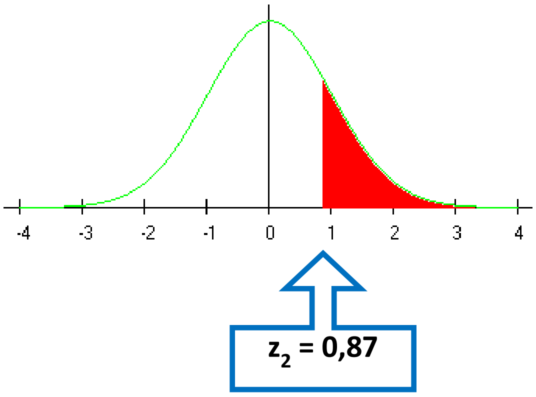
Área buscada = 0,0764



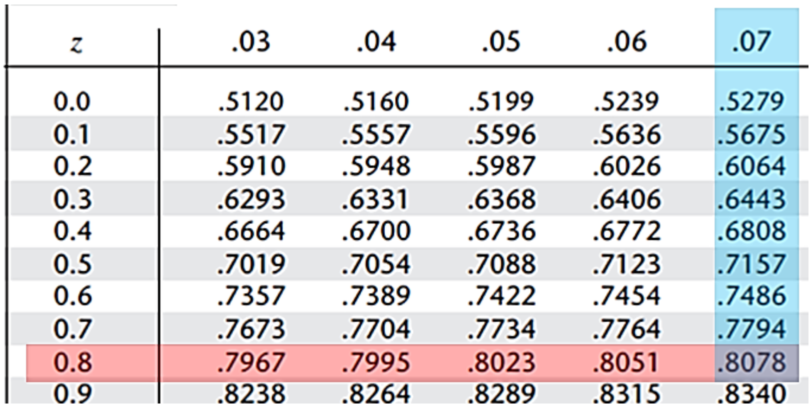
z	.00	.01	.02	.03	.04
−3.50 y menores	.0001				
−3.4	.0003	.0003	.0003	.0003	.0003
−3.3	.0005	.0005	.0005	.0004	.0004
−3.2	.0007	.0007	.0006	.0006	.0006
−3.1	.0010	.0009	.0009	.0009	.0008
−3.0	.0013	.0013	.0013	.0012	.0012
−2.9	.0019	.0018	.0018	.0017	.0016
−2.8	.0026	.0025	.0024	.0023	.0023
−2.7	.0035	.0034	.0033	.0032	.0031
−2.6	.0047	.0045	.0044	.0043	.0041
−2.5	.0062	.0060	.0059	.0057	.0055
−2.4	.0082	.0080	.0078	.0075	.0073
−2.3	.0107	.0104	.0102	.0099	.0096
−2.2	.0139	.0136	.0132	.0129	.0125
−2.1	.0179	.0174	.0170	.0166	.0162
−2.0	.0228	.0222	.0217	.0212	.0207
−1.9	.0287	.0281	.0274	.0268	.0262
−1.8	.0359	.0351	.0344	.0336	.0329
−1.7	.0446	.0436	.0427	.0418	.0409
−1.6	.0548	.0537	.0526	.0516	.0505
−1.5	.0668	.0655	.0643	.0630	.0618
−1.4	.0808	.0793	.0778	.0764	.0749
−1.3	.0968	.0951	.0934	.0918	.0901

Segundo caso: Área a la derecha de "z" (se calcula por complementario)

Calcule el área bajo la curva normal a la derecha de z_2 = 0,87

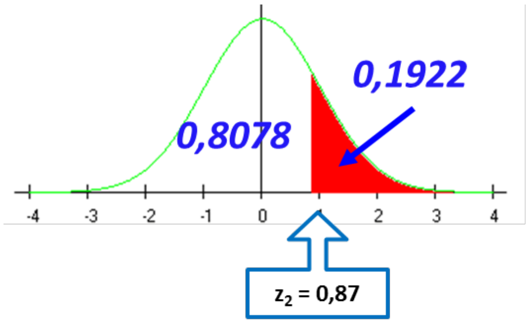


Para calcular dicha área se busca en la tabla "Z POSITIVA", ya que el valor de "z" es positivo. Luego se busca el valor 0,8 en la columna de "z" y el valor 7 en la fila superior como .07



z	.03	.04	.05	.06	.07
0.0	.5120	.5160	.5199	.5239	.5279
0.1	.5517	.5557	.5596	.5636	.5675
0.2	.5910	.5948	.5987	.6026	.6064
0.3	.6293	.6331	.6368	.6406	.6443
0.4	.6664	.6700	.6736	.6772	.6808
0.5	.7019	.7054	.7088	.7123	.7157
0.6	.7357	.7389	.7422	.7454	.7486
0.7	.7673	.7704	.7734	.7764	.7794
0.8	.7967	.7995	.8023	.8051	.8078
0.9	.8238	.8264	.8289	.8315	.8340

El área buscada se encuentra restando de 1 el área en la intersección de la fila con la columna. Es decir, se calcula el complementario de 0.8078. Esto se hace de esa manera ya que las tablas con que trabajamos sólo dan áreas a la izquierda. Haciendo uso de la propiedad que dice que el área bajo la curva normal estándar es igual a 1, entonces se tiene:

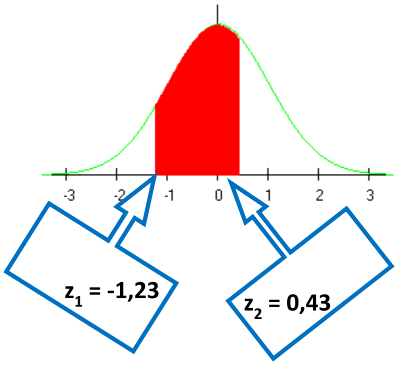


Por lo tanto:

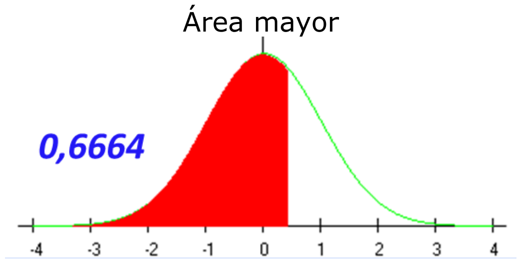
Área buscada = 1-0,8078 = 0,1922

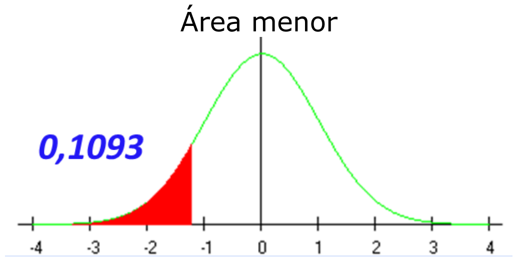
Tercer caso: Área entre dos valores "z" (se calcula por diferencia de áreas)

Calcule el área bajo la curva normal comprendida entre z_1 = -1,23 y z_2 = 0,43



Para calcular dicha área se busca en la tabla las áreas a la izquierda de ambos valores "z". Luego se resta el área menor del área mayor.


Área mayor


Área menor

Por lo tanto:

Área buscada = 0,6664 – 0,1093 = 0,5571

5. APLICACIONES DE LA DISTRIBUCIÓN NORMAL

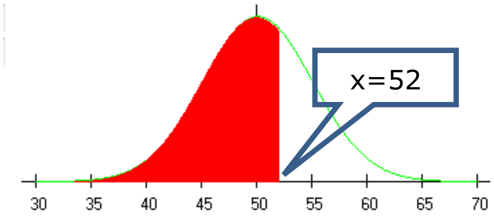
Ejemplo 1:

Los trabajadores de una empresa industrial tienen una edad media de 50 años y desviación estándar de 5 años. Se sabe que las edades de los trabajadores se distribuyen normalmente. Se pide:

- ¿Cuál es la probabilidad de elegir un trabajador que tenga menos de 52 años?

Solución:

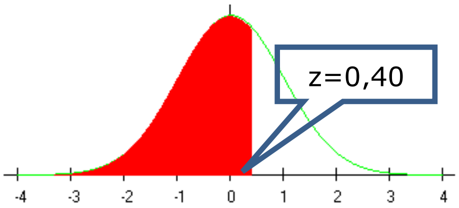
El primer paso será elaborar el gráfico que represente los datos. El valor de la media se ubica al centro de la figura en forma de campana. Luego se ubica el valor 52 y se pinta el área de la izquierda, ya que nos piden la probabilidad de edades menores a 52 años.



Ya que se observa que μ = 50 años y σ = 5 años no son valores que correspondan a una distribución normal estándar (donde μ = 0 y s = 1), es necesario realizar la estandarización, mediante el cálculo de la puntuación "z" equivalente a la edad de 52 años. Con este proceso habremos pasado de una distribución normal a una normal estándar.

Estandarizando: $z = \dfrac{52-50}{5} = 0,40$

Luego el ejercicio queda transformado a un caso de distribución normal estándar, como se aprecia en el gráfico:



En este punto podemos emplear las tablas de áreas bajo la curva normal estándar para calcular la probabilidad pedida.

Área pedida = 0,6554

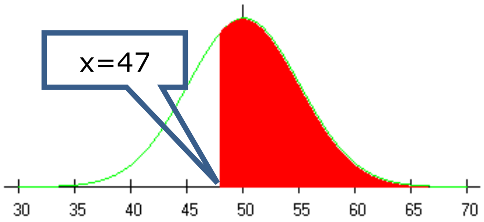
El área que acabamos de calcular es a la vez la probabilidad de elegir a un trabajador que tenga menos de 52 años.

Por lo tanto, la probabilidad de elegir a un trabajador que tenga menos de 52 años es de 0,6554.

- ¿Cuál es la probabilidad de elegir un trabajador cuya edad sea mayor a 47 años?

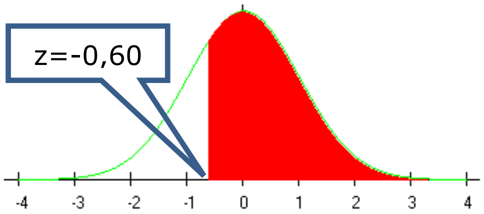
Solución

Graficamos nuestros datos:



Igual que en el caso anterior, estandarizamos el valor 47:

$$z = \frac{47-50}{5} = -0,60$$



En este punto podemos emplear las tablas de áreas bajo la curva normal estándar para calcular la probabilidad pedida.

Área pedida = 1–0,2743 = 0,7257

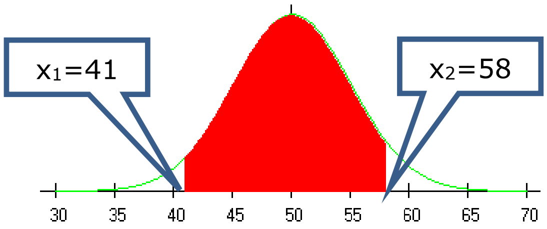
El área que acabamos de calcular es a la vez la probabilidad de elegir a un trabajador que tenga más de 47 años.

Por lo tanto, la probabilidad de elegir a un trabajador que tenga más de 47 años es de 0,7257.

· ¿Cuál es la probabilidad de que elegir un trabajador cuya edad esté comprendida entre 41 y 58 años?

Solución

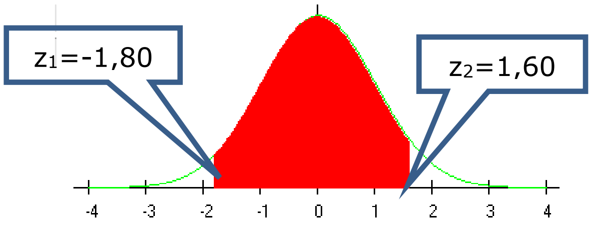
Graficamos nuestros datos:



Igual que en el caso anterior, estandarizamos el valor 41 y el 58:

$$z_1 = \frac{41-50}{5} = -1,80$$

$$z_2 = \frac{58-50}{5} = 1,60$$



En este punto podemos emplear las tablas de áreas bajo la curva normal estándar para calcular la probabilidad pedida.

Área pedida = 0,9452-0,0359 = 0,9093

El área que acabamos de calcular es a la vez la probabilidad de elegir a un trabajador cuya edad esté comprendida entre 41 y 58 años.

Por lo tanto, la probabilidad de elegir a un trabajador cuya edad esté comprendida entre 41 y 58 años es de 0,9093.

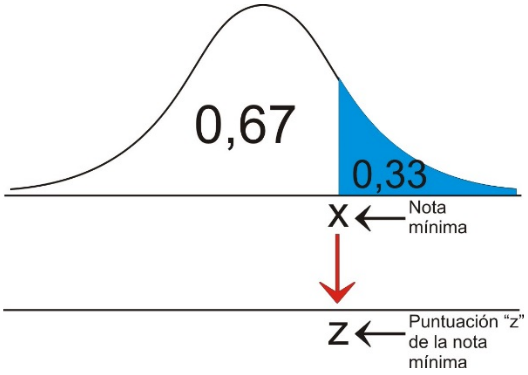
Ejemplo 2:

Cierta compañía convoca a un concurso para cubrir puestos de trabajo, considerando el siguiente criterio: Solamente aceptará al 33% de los postulantes que obtengan las más altas notas. Se conoce por experiencias anteriores que la nota media de los exámenes es de 11, con una desviación estándar de 2 y que las notas tienen una distribución normal.

- ¿Cuál es la nota mínima para cubrir un puesto de trabajo en dicha compañía?

Solución

Graficamos nuestros datos:



En primer lugar, determinamos cuál es el valor de "z" que tiene a su izquierda un área de 0,67 (lo que es lo mismo que buscar el valor "z" que tiene a su derecha un área de 0,33). Observando en la tabla vemos que corresponde a z = 0,44

Luego en la fórmula de estandarización reemplazamos:

$$z = \frac{x - Media}{Desv.Est.}$$

$$0,44 = \frac{x - 11}{2}$$

$$x = 11,88$$

Por lo tanto, la nota mínima para ingresar a dicha compañía es de 11,88 puntos.

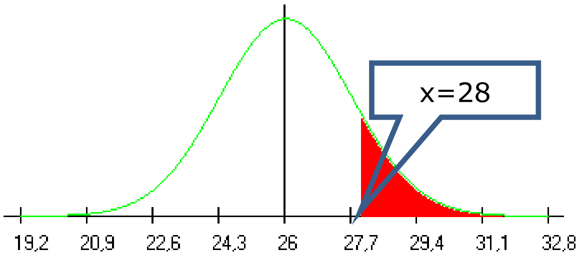
Ejemplo 3:

Una compañía paga a sus obreros un salario promedio de S/. 26 por hora, con una desviación estándar de S/. 1,7. Si los salarios tienen una distribución aproximadamente normal y se sabe además que actualmente cuentan con 220 obreros:

- ¿Cuántos obreros ganan salarios diarios mayores de S/. 28?

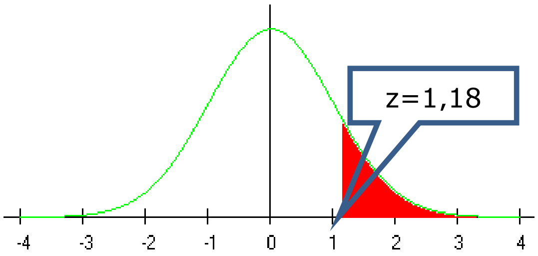
Solución:

Graficamos los datos:



Estandarizamos el valor 28 y lo llevamos a una distribución normal estándar:

$$z = \frac{28 - 26}{1,7} = 1,18$$



Calculamos el área restando de 1 el área a la izquierda de z = 1,18:

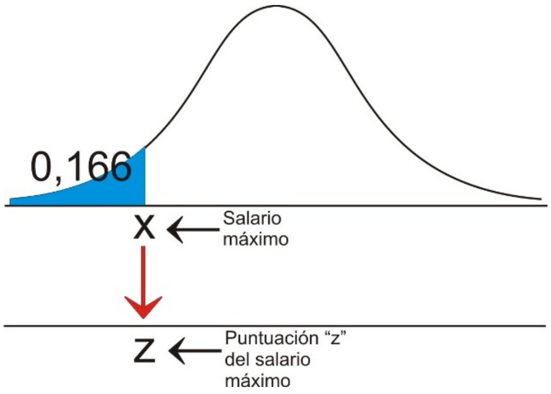
Área pedida = 1 – 0,8810 = 0,119

Luego multiplicamos el área por 100 y obtenemos 11,9%. Por lo tanto, el 11,9% de obreros tienen salarios mayores a S/. 28. Para determinar el número de obreros calculamos el 11,9% de 220, lo cual es 26,18. Pero este valor lo redondeamos a entero por tratarse de número de obreros, obteniendo finalmente 26 obreros.

- ¿Cuánto perciben como máximo el 16,6% de obreros que perciben los menores salarios?

Solución

Graficamos nuestros datos



Ahora buscamos el valor de "z" que a su izquierda tiene un área de 0,166. Mirando la tabla obtenemos z = -0,97

Luego reemplazamos en la fórmula de estandarización:

$$z = \frac{x - Media}{Desv.Est.}$$

$$-0,97 = \frac{x - 26}{1,7}$$

$$x = 24,35$$

Por lo tanto, el 16,6% de obreros cuyos salarios son los más bajos perciben como máximo S/. 24,35

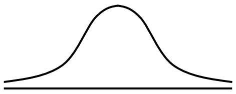
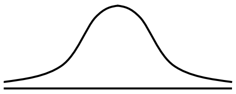
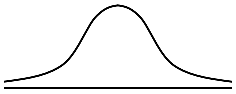
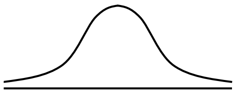
ACTIVIDAD FORMATIVA N.º 3

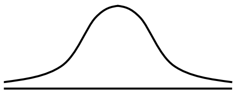
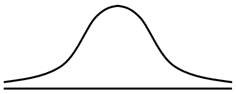
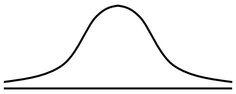
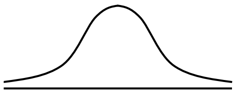
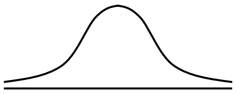
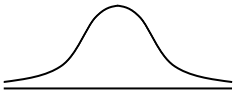
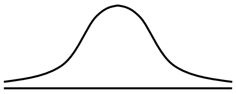
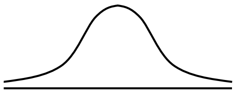
Identifica la distribución de probabilidad continua de un experimento aleatorio y calcula e interpreta la probabilidad de distribuciones normales.

INSTRUCCIONES

Resuelve los siguientes ejercicios e interpreta las probabilidades que calcules. No olvides que la interpretación es muy importante en nuestra asignatura.

1. ¿Qué requisitos son necesarios para que una distribución de probabilidad normal sea una distribución de probabilidad normal estándar?
2. Si usted determina que para la gráfica de una distribución normal estándar el área a la izquierda de una puntuación z es 0,4; ¿cuál es el área acumulativa que se localiza a la derecha de esa puntuación "z"?
3. Estandarizar los siguientes valores de una variable aleatoria continua:
   - x = 34 Media = 32 Desv. Est. = 2
   - x = 26 Media = 24 Desv. Est. = 5,1
   - x = 12 Media = 13 Desv. Est. = 0,25
   - x = 45 Media = 45 Desv. Est. = 3,65
   - x = 62 Media = 60,5 Desv. Est. = 3,56
4. Calcular el área debajo de la curva normal que se indica:

A la izquierda de z=-1,45 	A la izquierda de z=0,57 
A la izquierda de z=-2,34 	A la izquierda de z=2,87 

A la derecha de $z=-1,19$ 	A la derecha de $z=1,72$ 
A la derecha de $z=0,24$ 	A la derecha de $z=2,67$ 
Entre $z=-1,38$ y $z=1,64$ 	Entre $z=-2,03$ y $z=-0,42$ 
Entre $z=0,58$ y $z=1,94$ 	Entre $z=-2,41$ y $z=2,41$ 

4. Calcule el área acumulada correspondiente a:
   - $z < -1,36$
   - $z < -2,38$
   - $z < 1,76$
   - $z < 2,13$

5. Calcule el área acumulada correspondiente a:
   - $z > -1,25$
   - $z > 0,47$
   - $z > 2,35$
   - $z > –0,68$

6. Calcule el área acumulada correspondiente a:
   - $-1,24 < z < –0,36$
   - $-0,56 < z < 1,38$
   - $1,54 < z < 2,87$
   - $0 < z < 2,88$

7. Dado que z es la variable normal estándar, encuentre “z” en cada una de las situaciones siguientes:

   - El área a la izquierda de z es 0,9750
   - El área a la izquierda de z es 0,7291
   - El área a la derecha de z es 0,1314
   - El área a la derecha de z es 0,508
   - El área a la derecha de z es 0,3300
   - El área a la izquierda de z es 0,2119

8. Los trabajadores de una empresa industrial tienen una edad media de 51 años y desviación estándar de 5 años. Se pide:

   a) ¿Cuál es el porcentaje de trabajadores cuyas edades están entre 50 y 52,5 años?

   b) ¿Cuál es la probabilidad de que los trabajadores no sean mayores de 45 años?

   c) ¿Cuál es la probabilidad de que al elegir un trabajador su edad esté comprendida entre 41 y 58 años?

   d) ¿Cuál es la probabilidad de que los trabajadores tengan entre 55 y 60 años?

   e) ¿Cuál es la edad bajo la cual se encuentra el 20% de los trabajadores?

9. La estatura de los estudiantes de una cierta universidad está distribuida normalmente con una media de 69 pulgadas y una desviación estándar de 2 pulgadas. Se pide:

   a) ¿Cuál es la probabilidad de que un estudiantes mida más de 72 pulgadas?

   b) ¿Cuál es el porcentaje de estudiantes cuyas estaturas están entre 69 y 73 pulgadas?

   c) Si para ser seleccionado para integrar la selección de básquet, un estudiante debe estar en el 20% de los de mayor estatura. ¿Cuál es la estatura mínima para integrar dicha selección?

10. Un rodaje es considerado defectuoso, y por lo tanto es rechazado, si su diámetro es mayor que 2,02 pulgadas o menor que 1,98 pulgadas. ¿Cuál es la probabilidad de que un rodaje sea rechazado, si los diámetros de una partida de 10 000 rodajes están distribuidos normalmente con una media de 2 pulgadas y una desviación estándar de 0,01 pulgadas?

11. La media de los pesos de 500 estudiantes de un colegio es 70 kg y la desviación típica 3 kg. Suponiendo que los pesos se distribuyen normalmente, hallar cuántos estudiantes pesan:

    - Entre 60 kg y 75 kg.
    - Más de 90 kg.
    - 64 kg o menos.

12. La media de los sueldos de 600 empleados de una empresa es de 430 dólares y la desviación estándar o típica 40 dólares. Suponiendo que los sueldos se distribuyen normalmente, hallar cuántos empleados tienen sueldos

    a) Entre 350 y 450 dólares,

    b) De 500 a más dólares

13. En una muestra de estudiantes de Psicología, se encuentra que la nota promedio en Estadística fue 12 puntos, con una desviación estándar igual a 2. ¿Cuál es la probabilidad de que un alumno elegido al azar tenga nota entre 11 y 14?

14. ¿Cuál es la probabilidad de elegir a un ingeniero que tenga un ingreso económico mensual menor a 4700 nuevos soles, si la población esta normalmente distribuida y su media es S/. 4850 y la desviación es S/. 310?

15. ¿Cuál es la probabilidad de elegir aleatoriamente a una persona con menos de 30 años de edad en una distribución normal cuya media es de 25 años y una deviación estándar de 6 años?

16. El coeficiente intelectual (CI) de una muestra de profesionales se relaciona con una distribución normal donde $\mu=100$ puntos y $\sigma=10$ puntos.

    - ¿Qué calificación corresponde al percentil 25avo?
    - ¿Qué calificación corresponde al percentil 90avo?
    - Hallar el porcentaje de casos que tiene una calificación mayor a 91.

17. Se efectuó una investigación para saber durante cuántas horas de clase los estudiantes de una universidad (de 1500 estudiantes) navegan en internet diariamente. Los resultados se encontraron normalmente distribuidos con media =2,3 horas y Desv. Est. =0,7 horas. ¿Cuál es el porcentaje y número de casos que navegan en internet:

    - Más de dos horas diarias.
    - Menos de tres horas diarias.
    - Entre dos y tres horas diarias

18. La media de la cantidad de calorías (en cientos) que consumen diariamente los niños de una comunidad de Huancayo es 14,07 y la s=6,1. Si la población está compuesta por 950 niños y niñas y están normalmente distribuidos, halle:

    - La probabilidad de elegir a un niño o niña que consume menos de 18% de calorías diarias.
    - La probabilidad de elegir aleatoriamente un niño o niña que consume más del 14%

19. En una distribución normal el 68,79% de los miembros de una población tienen un puntaje z menor que cierto puntaje z. Determine el valor de éste último.

20. El número de horas diarias que una muestra de estudiantes universitarios ingresan diariamente en las redes sociales se distribuye normalmente donde $\mu=1{,}2$ horas y $\sigma=0{,}53$ horas.

   - ¿Qué calificación z corresponde al percentil 30avo?
   - ¿Qué calificación corresponde al percentil 80avo?
   - Hallar el porcentaje de casos que tiene una calificación mayor a 1 hora.

VIDEOS

Distribución normal: Conceptos y propiedades

https://www.youtube.com/watch?v=b_Oee84PrGg

LECTURA SELECCIONADA N.º 1

¿Qué es una distribución de probabilidad?

Estadística para Administración y Economía.

Richard Levin. Pág. 128

¿Qué es una distribución de probabilidad?

En el capítulo 2 describimos a las distribuciones de frecuencias como una forma útil de resumir las variaciones en los datos observados. Preparamos distribuciones de frecuencias haciendo una lista de todos los resultados posibles de un experimento para después indicar la frecuencia observada de cada resultado posible. Las *distribuciones de probabilidad* están relacionadas con las distribuciones de frecuencias. **De hecho, podemos pensar que una distribución de probabilidad es una distribución de frecuencias teórica.** ¿Qué significa lo anterior? Una distribución de frecuencias teórica es una distribución de probabilidades que describe la forma en que se *espera* varíen los resultados. Como estas distribuciones representan expectativas de que algo suceda, resultan modelos útiles para hacer inferencias y tomar decisiones en condiciones de incertidumbre. En capítulos posteriores, analizaremos los métodos que utilizamos bajo tales condiciones.

Ejemplos de distribuciones de probabilidad

Para comenzar nuestro estudio de las distribuciones de probabilidad, regresemos a la idea de la moneda no alterada que introdujimos en el capítulo 4. Suponga que lanzamos esa moneda dos veces.

La tabla 5-1 lista los posibles resultados para este experimento de dos lanzamientos. [Cara (*head*) está representada con una H; cruz (*tail*) con una T.] Suponga ahora que nos interesa formular una distribución de probabilidad del número de cruces (T) que podrían caer cuando lanzamos la moneda dos veces. Podríamos empezar por anotar cualquier resultado que *no* contenga cruces. Con una moneda no alterada, estaríamos hablando exclusivamente del tercer resultado de la tabla 5-1: *H, H*. Luego anotaríamos los resultados que tuvieran sólo una cruz (segundo y cuarto resultados de la tabla 5-1) y, por último, incluiríamos el primer resultado que contiene dos cruces. En la tabla 5-2 ordenamos los resultados de la 5-1 para enfatizar el número de cruces contenidas en cada resultado. En este punto, debemos tener cuidado y considerar que la tabla 5-2 *no* representa el resultado real de lanzar una moneda no alterada dos veces. Más bien, se trata del resultado *teórico*, es decir, representa la forma en que *esperamos* que se comporte nuestro experimento de dos lanzamientos.

Tabla 5-1

Posibles resultados de lanzar dos veces una moneda no alterada

Primer lanzamiento	Segundo lanzamiento	Número de cruces en dos lanzamientos	Probabilidad de los cuatro resultados posibles
T	*T*	2	0.5 × 0.5 = 0.25
T	*H*	1	0.5 × 0.5 = 0.25
H	*H*	0	0.5 × 0.5 = 0.25
H	*T*	1	0.5 × 0.5 = 0.25
			1.00

Tabla 5-2

Distribución de probabilidad del número posible de cruces que se obtienen al lanzar dos veces una moneda no alterada

Número de cruces (T)	Lanzamientos	Probabilidad de este resultado, P(*T*)
0	(*H*, *H*)	0.25
1	(*T*, *H*) + (*H*, *T*)	0.50
2	(*T*, *T*)	0.25

Podemos representar gráficamente la distribución de probabilidad de la tabla 5-2. Para ello, colocamos en una gráfica el número de cruces que deberíamos ver en dos lanzamientos contra la probabilidad de que este número se presente. La figura 5-1 ilustra este resultado.

Analicemos otro ejemplo. Una candidata política para un puesto en el gobierno local está considerando los votos que puede obtener en las elecciones que se avecinan. Suponga que los votos pueden tomar sólo cuatro valores posibles. Si la estimación de la candidata es como sigue:

Número de votos	1,000	2,000	3,000	4,000	
Probabilidad de que éstos se obtengan	0.1	0.3	0.4	0.2	**Total 1.0**

Entonces la gráfica de la distribución de probabilidad que representa sus expectativas debe ser como la que mostramos en la figura 5-2. Antes de analizar otros aspectos de las distribuciones de probabilidad, debemos señalar que **una distribución de frecuencias es un listado de las frecuencias observadas de todos los resultados de un experimento que se presentaron realmente cuando se efectuó éste, mientras que una distribución de probabilidad es un listado de las probabilidades de todos los posibles resultados que *podrían* obtenerse si el experimento se *llevara* a cabo.** También, como podemos darnos cuenta en los ejemplos de las figuras 5-1 y 5-2, las distribuciones de probabilidad pueden basarse en consideraciones teóricas (los lanzamientos de una moneda) o en una estimación subjetiva de la posibilidad de ciertos resultados (la estimación de la candidata). Las distribuciones de probabilidad se pueden basar también en la experiencia. Los actuarios de las compañías de seguros, por ejemplo, determinan las pólizas de seguros haciendo uso de los largos años de experiencia que las compañías tienen acerca de los índices de mortalidad para establecer la probabilidad de muerte entre los diferentes grupos de edad.

LECTURA SELECCIONADA No 2:

Yo mido 1,68. ¿Por qué la probabilidad de que una estatura

sea 1,68 calculada con la distribución Normal es 0?

Roberto Behar Gutiérrez, Pere Grima Cintas. Página 51

Se trata de una aparente contradicción debido a que en la práctica tratamos como discretas a las variables que en teoría consideramos como continuas.

Cuando decimos que la variable aleatoria X sigue una distribución N(μ; s), estamos diciendo también que X es una variable continua que puede tomar infinitos valores.

Sabemos que la probabilidad de que tome valores comprendidos entre a y b es igual al área definida por su función densidad de probabilidad entre estos valores, es decir:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Por tanto

$$P(X = 1{,}68) = P(1{,}68 \leq X \leq 1{,}68) = \int_{1{,}68}^{1{,}68} f(x)dx = 0$$

Pero cuando decimos que alguien mide 1,68 metros no queremos decir que mide

exactamente 1,680000000… con infinitos ceros, sino que entendemos que el valor se ha redondeado, y lo que queremos decir es que su altura está comprendida entre 1,675 y 1,685 (si fuera menor de 1,675 lo habríamos aproximado a 1,67 y si mayor de 1,685 a 1,69)

Suponiendo que las alturas se distribuyan según una Normal con media μ =1,70 m y desviación tipo s =0,07 m, la probabilidad de que una persona mida 1,68 m (entendido como valor redondeado, que es como hablamos), es de 0,0547. Tranquilo, usted existe.

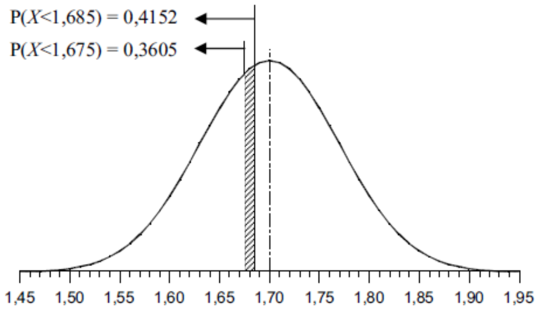


Figura 15.1. *Probabilidad de que X ~ N(1,70; 0,07) tome valores comprendidos entre 1,675 y 1,685*

Todas las variables continuas se discretizan para tratar con ellas en la práctica. Casi

siempre se redondea, como en el caso de las alturas, aunque en algunos casos se trunca, como hacemos con las edades. Si hoy es día de las elecciones generales y usted cumple los 18 años mañana, usted hoy tiene 17+364/365 = 17,997 años. Pero seguramente no le dejarán votar.

Observación final: Un argumento falaz, pero muchas veces convincente haciendo mención a aquello de los casos favorables partido por los casos posibles, es considerar que si la variable puede tomar infinitos valores, la probabilidad de que tome uno en concreto es cero. Esto es falso porque la regla de casos favorables partido por casos posibles solo vale cuando todos los sucesos tienen la misma probabilidad de ocurrir, y este no es nuestro caso. Por otra parte, es perfectamente posible que una variable pueda tomar infinitos valores y que ninguno de ellos tenga probabilidad cero. Piense en la distribución de Poisson, por ejemplo.

PRUEBA DE DESARROLLO – UNIDAD IV

1. Distribución Binomial: El 48% de los habitantes que viven en un pueblo joven se dedicaban al comercio ambulatorio. Se seleccionan 8 pobladores al azar. Qué probabilidad hay de que exactamente 5 se dediquen al comercio ambulatorio

2. Distribución Poisson: La Financiera de Crédito "Edificar" recibe en promedio 15 solicitudes de préstamos para mejoramiento de vivienda por mes. ¿Cuál es la probabilidad de que reciba exactamente 14 solicitudes el siguiente mes?

3. Una población normal tiene una media de 20,00 y una desviación estándar de 4,00. Halle el valor "z" correspondiente a 25,00

4. Una empresa fabrica fluorescentes cuya duración es normalmente distribuida con una media igual a 900 horas y una desviación estándar de 73 horas. Calcule la probabilidad de que al elegir un fluorescente cualquiera éste dure entre 847 a 910 horas.

5. Un investigador concluye que las ratas viven un promedio de 42 meses. Suponiendo que las vidas de tales ratas siguen una distribución normal con una desviación estándar de 6,5 meses. ¿Cuál es la probabilidad de que una rata elegida al azar viva menos de 40 meses?

6. Los panes de cierta panadería tienen un peso promedio de 10 gramos y una desviación estándar de 2 gramos. Suponiendo que los pesos están normalmente distribuidos y que diariamente se producen 900 unidades de pan. ¿Cuántos panes pesan más de 12 gramos?

GLOSARIO DE LA UNIDAD IV

D

Distribución binomial: Distribución donde los sucesos tienen la misma probabilidad, con la presencia de éxito y fracaso y en un número fijo de ensayos.

Distribución de Poisson: Distribución de probabilidad discreta que se aplica a ocurrencias de algún suceso durante un intervalo de tiempo, distancia, área, volumen u otra unidad similar que se especifique.

Distribución de probabilidad: Conjunto de valores de una variable aleatoria junto con sus probabilidades correspondientes.

Distribución hipergeométrica: Se diferencia de la binomial en que la probabilidad de los sucesos no se mantiene constante.

Distribución muestral de proporciones: Distribución de probabilidad de las proporciones muestrales, donde todas las muestras tienen el mismo tamaño muestral n.

Distribución normal: Distribución de probabilidad con forma de campana, simétrica y asintótica por ambos extremos.

Distribución normal estándar: Distribución normal con una media igual a cero y una desviación estándar igual a 1

V

Valor crítico z: Número de veces que la diferencia entre la observación dada y la media contiene a la desviación estándar.

Valor esperado: Promedio ponderado de los resultados de un experimento.

Variable Aleatoria: Variable que toma diferentes valores como resultado de un experimento aleatorio.

BIBLIOGRAFÍA DE LA UNIDAD IV

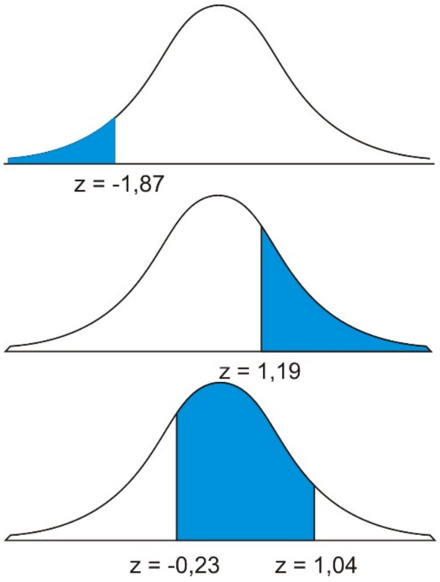
Devore, J. (2008). *Probabilidad y Estadística* (7 ed.). México: Editorial Cengage Learning Editores S.A.

Triola, M.(2009). *Estadística.* México: Pearson Educación.

Webster, A. (2000). *Estadística aplicada a los negocios y la economía.* (3 ed.). Colombia. Editorial Mc. Graw-Hill.

AUTOEVALUACION No 4

1. Calcule el área sombreada en cada uno de los casos.



2. Una muestra de los tiempos empleados por un grupo de estudiantes para resolver una prueba de Estadística tiene una media de 55 minutos con una desviación estándar de 6,89 minutos. La señorita Paredes ha terminado la prueba en 58 minutos. Calcule la puntuación "z" para este tiempo.

3. El promedio de los pesos de 60 niños de un cierto hospital es de 10,5 kg con una desviación estándar o típica de 1,5 kg. Suponiendo que los pesos se distribuyen normalmente, calcule la probabilidad de que un niño elegido al azar pese entre 9 kg y 11 kg.

4. En cierta comunidad, el gasto promedio de alimentación diaria por familia es de S/. 9,80, con una desviación estándar de S/. 0,53. Si los gastos tienen una distribución normal, ¿qué porcentaje de familias gasta menos de S/. 9 de su ingreso?

5. Las estaturas de 1000 reclutas del ejército están distribuidas normalmente con una media de 1,65 m y una desviación estándar de 0,14 m. ¿Cuántos reclutas tendrán estaturas comprendidas entre 1,63 m y 1,68 m?

ANEXO "A"

Tablas de distribución normal

TABLA A-1 Probabilidades binomiales

							p							
n	*x*	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	*x*
2	0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	0+	0
	1	.020	.095	.180	.320	.420	.480	.500	.480	.420	.320	.180	.095	.020	1
	2	0+	.002	.010	.040	.090	.160	.250	.360	.490	.640	.810	.902	.980	2
3	0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	0+	0+	0
	1	.029	.135	.243	.384	.441	.432	.375	.288	.189	.096	.027	.007	0+	1
	2	0+	.007	.027	.096	.189	.288	.375	.432	.441	.384	.243	.135	.029	2
	3	0+	0+	.001	.008	.027	.064	.125	.216	.343	.512	.729	.857	.970	3
4	0	.961	.815	.656	.410	.240	.130	.062	.026	.008	.002	0+	0+	0+	0
	1	.039	.171	.292	.410	.412	.346	.250	.154	.076	.026	.004	0+	0+	1
	2	.001	.014	.049	.154	.265	.346	.375	.346	.265	.154	.049	.014	.001	2
	3	0+	0+	.004	.026	.076	.154	.250	.346	.412	.410	.292	.171	.039	3
	4	0+	0+	0+	.002	.008	.026	.062	.130	.240	.410	.656	.815	.961	4
5	0	.951	.774	.590	.328	.168	.078	.031	.010	.002	0+	0+	0+	0+	0
	1	.048	.204	.328	.410	.360	.259	.156	.077	.028	.006	0+	0+	0+	1
	2	.001	.021	.073	.205	.309	.346	.312	.230	.132	.051	.008	.001	0+	2
	3	0+	.001	.008	.051	.132	.230	.312	.346	.309	.205	.073	.021	.001	3
	4	0+	0+	0+	.006	.028	.077	.156	.259	.360	.410	.328	.204	.048	4
	5	0+	0+	0+	0+	.002	.010	.031	.078	.168	.328	.590	.774	.951	5
6	0	.941	.735	.531	.262	.118	.047	.016	.004	.001	0+	0+	0+	0+	0
	1	.057	.232	.354	.393	.303	.187	.094	.037	.010	.002	0+	0+	0+	1
	2	.001	.031	.098	.246	.324	.311	.234	.138	.060	.015	.001	0+	0+	2
	3	0+	.002	.015	.082	.185	.276	.312	.276	.185	.082	.015	.002	0+	3
	4	0+	0+	.001	.015	.060	.138	.234	.311	.324	.246	.098	.031	.001	4
	5	0+	0+	0+	.002	.010	.037	.094	.187	.303	.393	.354	.232	.057	5
	6	0+	0+	0+	0+	.001	.004	.016	.047	.118	.262	.531	.735	.941	6
7	0	.932	.698	.478	.210	.082	.028	.008	.002	0+	0+	0+	0+	0+	0
	1	.066	.257	.372	.367	.247	.131	.055	.017	.004	0+	0+	0+	0+	1
	2	.002	.041	.124	.275	.318	.261	.164	.077	.025	.004	0+	0+	0+	2
	3	0+	.004	.023	.115	.227	.290	.273	.194	.097	.029	.003	0+	0+	3
	4	0+	0+	.003	.029	.097	.194	.273	.290	.227	.115	.023	.004	0+	4
	5	0+	0+	0+	.004	.025	.077	.164	.261	.318	.275	.124	.041	.002	5
	6	0+	0+	0+	0+	.004	.017	.055	.131	.247	.367	.372	.257	.066	6
	7	0+	0+	0+	0+	0+	.002	.008	.028	.082	.210	.478	.698	.932	7
8	0	.923	.663	.430	.168	.058	.017	.004	.001	0+	0+	0+	0+	0+	0
	1	.075	.279	.383	.336	.198	.090	.031	.008	.001	0+	0+	0+	0+	1
	2	.003	.051	.149	.294	.296	.209	.109	.041	.010	.001	0+	0+	0+	2
	3	0+	.005	.033	.147	.254	.279	.219	.124	.047	.009	0+	0+	0+	3
	4	0+	0+	.005	.046	.136	.232	.273	.232	.136	.046	.005	0+	0+	4
	5	0+	0+	0+	.009	.047	.124	.219	.279	.254	.147	.033	.005	0+	5
	6	0+	0+	0+	.001	.010	.041	.109	.209	.296	.294	.149	.051	.003	6
	7	0+	0+	0+	0+	.001	.008	.031	.090	.198	.336	.383	.279	.075	7
	8	0+	0+	0+	0+	0+	.001	.004	.017	.058	.168	.430	.663	.923	8

NOTA: 0+ representa una probabilidad positiva menor que 0.0005. *(continúa)*

Tomado del libro de Estadística de Mario Triola, décima edición pp. 731

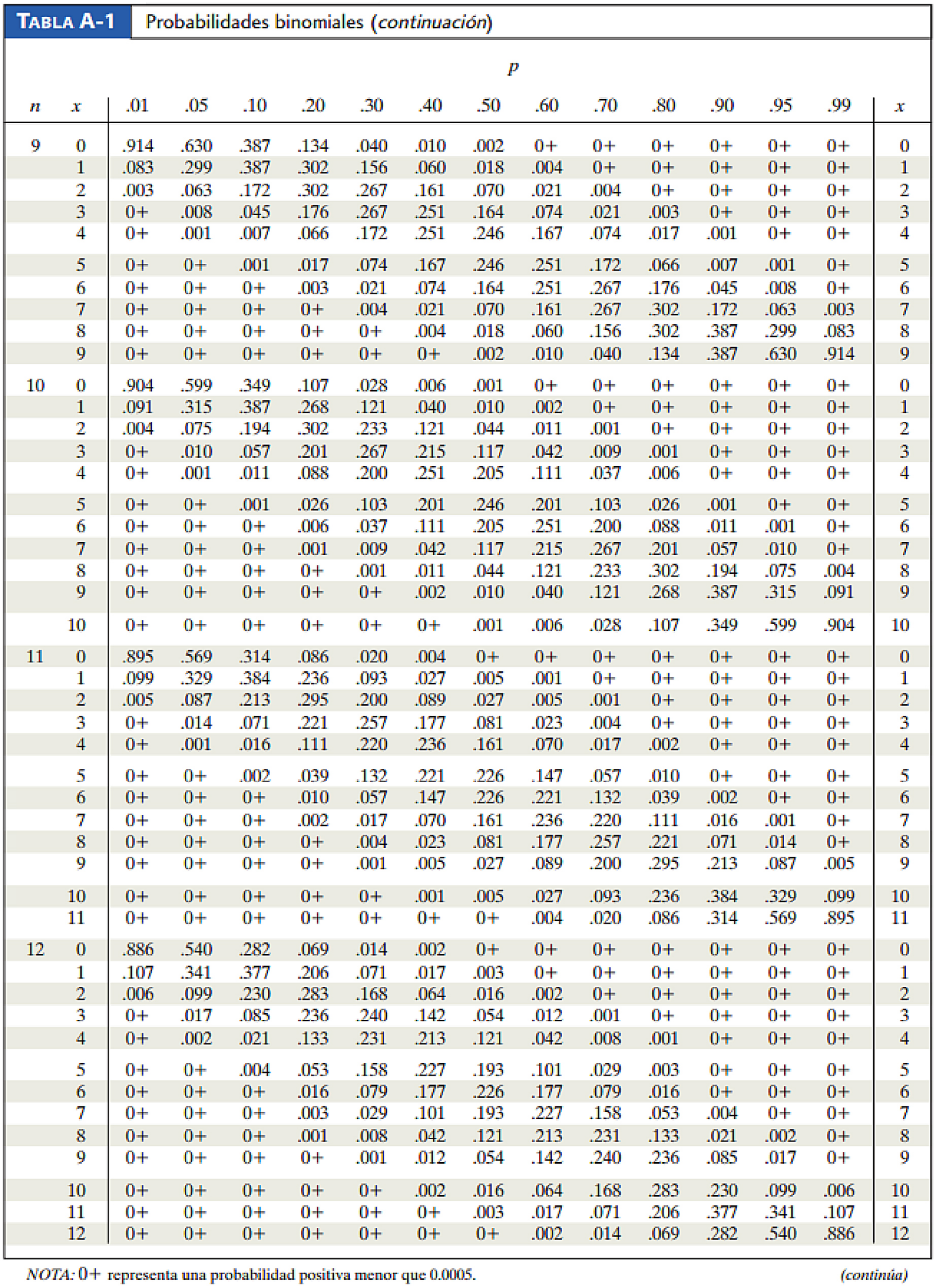


TABLA A-1 Probabilidades binomiales (*continuación*)

		p													
n	*x*	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	*x*
9	0	.914	.630	.387	.134	.040	.010	.002	0+	0+	0+	0+	0+	0+	0
	1	.083	.299	.387	.302	.156	.060	.018	.004	0+	0+	0+	0+	0+	1
	2	.003	.063	.172	.302	.267	.161	.070	.021	.004	0+	0+	0+	0+	2
	3	0+	.008	.045	.176	.267	.251	.164	.074	.021	.003	0+	0+	0+	3
	4	0+	.001	.007	.066	.172	.251	.246	.167	.074	.017	.001	0+	0+	4
	5	0+	0+	.001	.017	.074	.167	.246	.251	.172	.066	.007	.001	0+	5
	6	0+	0+	0+	.003	.021	.074	.164	.251	.267	.176	.045	.008	0+	6
	7	0+	0+	0+	0+	.004	.021	.070	.161	.267	.302	.172	.063	.003	7
	8	0+	0+	0+	0+	0+	.004	.018	.060	.156	.302	.387	.299	.083	8
	9	0+	0+	0+	0+	0+	0+	.002	.010	.040	.134	.387	.630	.914	9
10	0	.904	.599	.349	.107	.028	.006	.001	0+	0+	0+	0+	0+	0+	0
	1	.091	.315	.387	.268	.121	.040	.010	.002	0+	0+	0+	0+	0+	1
	2	.004	.075	.194	.302	.233	.121	.044	.011	.001	0+	0+	0+	0+	2
	3	0+	.010	.057	.201	.267	.215	.117	.042	.009	.001	0+	0+	0+	3
	4	0+	.001	.011	.088	.200	.251	.205	.111	.037	.006	0+	0+	0+	4
	5	0+	0+	.001	.026	.103	.201	.246	.201	.103	.026	.001	0+	0+	5
	6	0+	0+	0+	.006	.037	.111	.205	.251	.200	.088	.011	.001	0+	6
	7	0+	0+	0+	.001	.009	.042	.117	.215	.267	.201	.057	.010	0+	7
	8	0+	0+	0+	0+	.001	.011	.044	.121	.233	.302	.194	.075	.004	8
	9	0+	0+	0+	0+	0+	.002	.010	.040	.121	.268	.387	.315	.091	9
	10	0+	0+	0+	0+	0+	0+	.001	.006	.028	.107	.349	.599	.904	10
11	0	.895	.569	.314	.086	.020	.004	0+	0+	0+	0+	0+	0+	0+	0
	1	.099	.329	.384	.236	.093	.027	.005	.001	0+	0+	0+	0+	0+	1
	2	.005	.087	.213	.295	.200	.089	.027	.005	.001	0+	0+	0+	0+	2
	3	0+	.014	.071	.221	.257	.177	.081	.023	.004	0+	0+	0+	0+	3
	4	0+	.001	.016	.111	.220	.236	.161	.070	.017	.002	0+	0+	0+	4
	5	0+	0+	.002	.039	.132	.221	.226	.147	.057	.010	0+	0+	0+	5
	6	0+	0+	0+	.010	.057	.147	.226	.221	.132	.039	.002	0+	0+	6
	7	0+	0+	0+	.002	.017	.070	.161	.236	.220	.111	.016	.001	0+	7
	8	0+	0+	0+	0+	.004	.023	.081	.177	.257	.221	.071	.014	0+	8
	9	0+	0+	0+	0+	.001	.005	.027	.089	.200	.295	.213	.087	.005	9
	10	0+	0+	0+	0+	0+	.001	.005	.027	.093	.236	.384	.329	.099	10
	11	0+	0+	0+	0+	0+	0+	0+	.004	.020	.086	.314	.569	.895	11
12	0	.886	.540	.282	.069	.014	.002	0+	0+	0+	0+	0+	0+	0+	0
	1	.107	.341	.377	.206	.071	.017	.003	0+	0+	0+	0+	0+	0+	1
	2	.006	.099	.230	.283	.168	.064	.016	.002	0+	0+	0+	0+	0+	2
	3	0+	.017	.085	.236	.240	.142	.054	.012	.001	0+	0+	0+	0+	3
	4	0+	.002	.021	.133	.231	.213	.121	.042	.008	.001	0+	0+	0+	4
	5	0+	0+	.004	.053	.158	.227	.193	.101	.029	.003	0+	0+	0+	5
	6	0+	0+	0+	.016	.079	.177	.226	.177	.079	.016	0+	0+	0+	6
	7	0+	0+	0+	.003	.029	.101	.193	.227	.158	.053	.004	0+	0+	7
	8	0+	0+	0+	.001	.008	.042	.121	.213	.231	.133	.021	.002	0+	8
	9	0+	0+	0+	0+	.001	.012	.054	.142	.240	.236	.085	.017	0+	9
	10	0+	0+	0+	0+	0+	.002	.016	.064	.168	.283	.230	.099	.006	10
	11	0+	0+	0+	0+	0+	0+	.003	.017	.071	.206	.377	.341	.107	11
	12	0+	0+	0+	0+	0+	0+	0+	.002	.014	.069	.282	.540	.886	12

NOTA: 0+ representa una probabilidad positiva menor que 0.0005. *(continúa)*

Tomado del libro de Estadística de Mario Triola, décima edición pp. 732

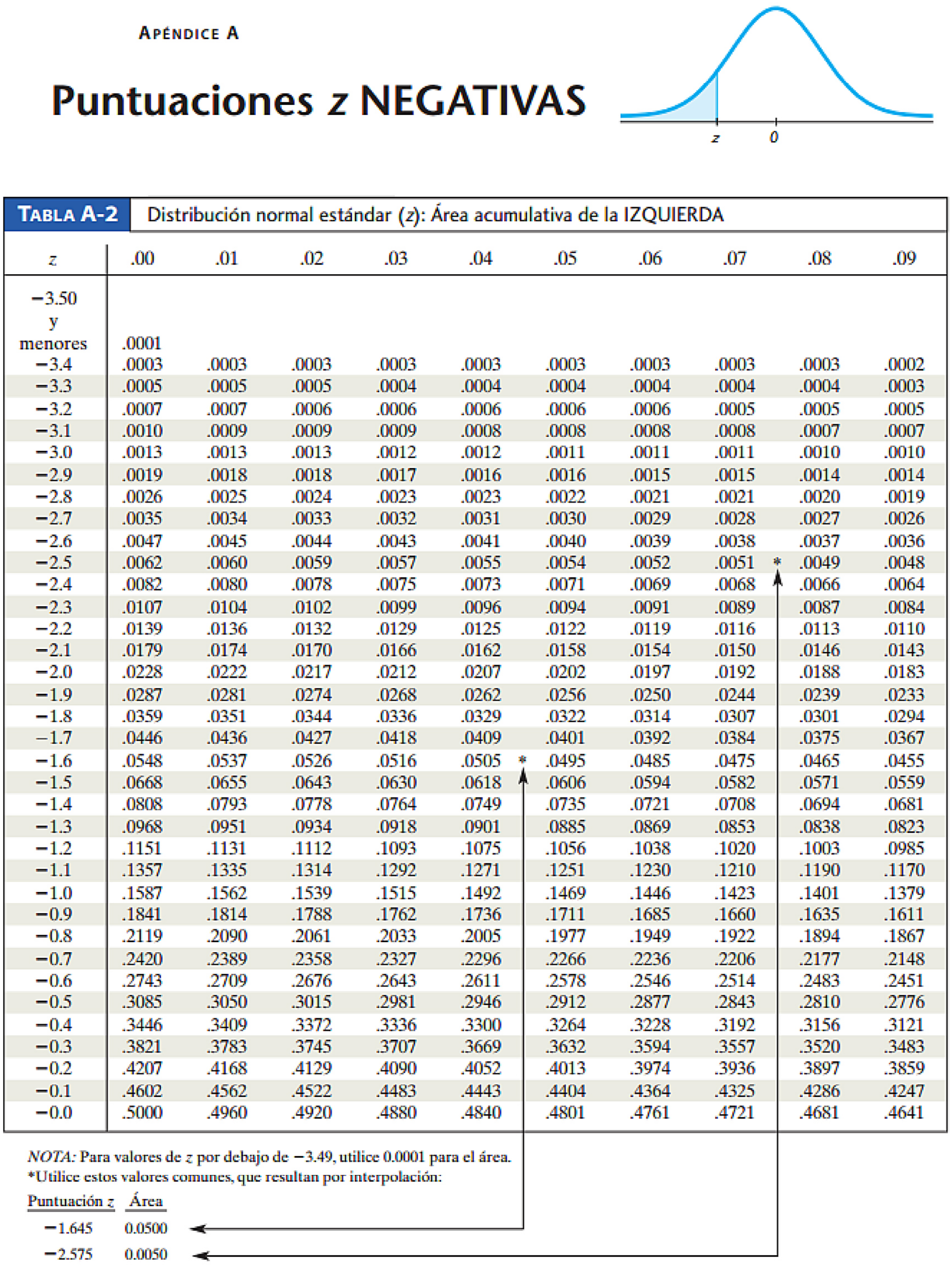
TABLA A-1 Probabilidades binomiales (*continuación*)

		p													
n	*x*	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	*x*
13	0	.878	.513	.254	.055	.010	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.115	.351	.367	.179	.054	.011	.002	0+	0+	0+	0+	0+	0+	1
	2	.007	.111	.245	.268	.139	.045	.010	.001	0+	0+	0+	0+	0+	2
	3	0+	.021	.100	.246	.218	.111	.035	.006	.001	0+	0+	0+	0+	3
	4	0+	.003	.028	.154	.234	.184	.087	.024	.003	0+	0+	0+	0+	4
	5	0+	0+	.006	.069	.180	.221	.157	.066	.014	.001	0+	0+	0+	5
	6	0+	0+	.001	.023	.103	.197	.209	.131	.044	.006	0+	0+	0+	6
	7	0+	0+	0+	.006	.044	.131	.209	.197	.103	.023	.001	0+	0+	7
	8	0+	0+	0+	.001	.014	.066	.157	.221	.180	.069	.006	0+	0+	8
	9	0+	0+	0+	0+	.003	.024	.087	.184	.234	.154	.028	.003	0+	9
	10	0+	0+	0+	0+	.001	.006	.035	.111	.218	.246	.100	.021	0+	10
	11	0+	0+	0+	0+	0+	.001	.010	.045	.139	.268	.245	.111	.007	11
	12	0+	0+	0+	0+	0+	0+	.002	.011	.054	.179	.367	.351	.115	12
	13	0+	0+	0+	0+	0+	0+	0+	.001	.010	.055	.254	.513	.878	13
14	0	.869	.488	.229	.044	.007	.001	0+	0+	0+	0+	0+	0+	0+	0
	1	.123	.359	.356	.154	.041	.007	.001	0+	0+	0+	0+	0+	0+	1
	2	.008	.123	.257	.250	.113	.032	.006	.001	0+	0+	0+	0+	0+	2
	3	0+	.026	.114	.250	.194	.085	.022	.003	0+	0+	0+	0+	0+	3
	4	0+	.004	.035	.172	.229	.155	.061	.014	.001	0+	0+	0+	0+	4
	5	0+	0+	.008	.086	.196	.207	.122	.041	.007	0+	0+	0+	0+	5
	6	0+	0+	.001	.032	.126	.207	.183	.092	.023	.002	0+	0+	0+	6
	7	0+	0+	0+	.009	.062	.157	.209	.157	.062	.009	0+	0+	0+	7
	8	0+	0+	0+	.002	.023	.092	.183	.207	.126	.032	.001	0+	0+	8
	9	0+	0+	0+	0+	.007	.041	.122	.207	.196	.086	.008	0+	0+	9
	10	0+	0+	0+	0+	.001	.014	.061	.155	.229	.172	.035	.004	0+	10
	11	0+	0+	0+	0+	0+	.003	.022	.085	.194	.250	.114	.026	0+	11
	12	0+	0+	0+	0+	0+	.001	.006	.032	.113	.250	.257	.123	.008	12
	13	0+	0+	0+	0+	0+	0+	.001	.007	.041	.154	.356	.359	.123	13
	14	0+	0+	0+	0+	0+	0+	0+	.001	.007	.044	.229	.488	.869	14
15	0	.860	.463	.206	.035	.005	0+	0+	0+	0+	0+	0+	0+	0+	0
	1	.130	.366	.343	.132	.031	.005	0+	0+	0+	0+	0+	0+	0+	1
	2	.009	.135	.267	.231	.092	.022	.003	0+	0+	0+	0+	0+	0+	2
	3	0+	.031	.129	.250	.170	.063	.014	.002	0+	0+	0+	0+	0+	3
	4	0+	.005	.043	.188	.219	.127	.042	.007	.001	0+	0+	0+	0+	4
	5	0+	.001	.010	.103	.206	.186	.092	.024	.003	0+	0+	0+	0+	5
	6	0+	0+	.002	.043	.147	.207	.153	.061	.012	.001	0+	0+	0+	6
	7	0+	0+	0+	.014	.081	.177	.196	.118	.035	.003	0+	0+	0+	7
	8	0+	0+	0+	.003	.035	.118	.196	.177	.081	.014	0+	0+	0+	8
	9	0+	0+	0+	.001	.012	.061	.153	.207	.147	.043	.002	0+	0+	9
	10	0+	0+	0+	0+	.003	.024	.092	.186	.206	.103	.010	.001	0+	10
	11	0+	0+	0+	0+	.001	.007	.042	.127	.219	.188	.043	.005	0+	11
	12	0+	0+	0+	0+	0+	.002	.014	.063	.170	.250	.129	.031	0+	12
	13	0+	0+	0+	0+	0+	0+	.003	.022	.092	.231	.267	.135	.009	13
	14	0+	0+	0+	0+	0+	0+	0+	.005	.031	.132	.343	.366	.130	14
	15	0+	0+	0+	0+	0+	0+	0+	0+	.005	.035	.206	.463	.860	15

NOTA: 0+ representa una probabilidad positiva menor que 0.0005.

De Frederick C. Mosteller, Robert E. K. Rourke y George B. Thomas Jr., *Probability with Statistical Applications*, 2a. ed., © 1970 Addison-Wesley Publishing Co., Reading, MA. Reimpreso bajo permiso.

Tomado del libro de Estadística de Mario Triola, décima edición pp. 733



APÉNDICE A

Puntuaciones *z* NEGATIVAS

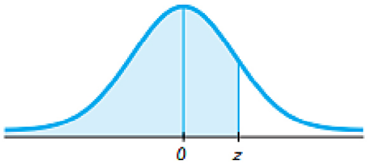
TABLA A-2 Distribución normal estándar (*z*): Área acumulativa de la IZQUIERDA

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 y menores	.0001									
−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

NOTA: Para valores de *z* por debajo de −3.49, utilice 0.0001 para el área.
*Utilice estos valores comunes, que resultan por interpolación:

Puntuación *z*	Área
−1.645	0.0500
−2.575	0.0050

Tomado del libro de Estadística de Mario Triola, décima edición pp. 734



Apéndice A

Puntuaciones *z* POSITIVAS

TABLA A-2 (*continuación*) Área acumulativa de la IZQUIERDA

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495 *	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949 *	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50 y mayores	.9999									

NOTA: Para valores de *z* por encima de 3.49, utilice 0.9999 para el área.
*Utilice estos valores comunes, que resultan por interpolación:

Puntuación *z*	Área
1.645	0.9500
2.575	0.9950

Valores comunes críticos

Nivel de confianza	Valor crítico
0.90	1.645
0.95	1.96
0.99	2.575

Tomado del libro de Estadística de Mario Triola, décima edición pp. 735

ANEXO Nº 1:
CLAVES DE RESPUESTAS A LAS AUTOEVALUACIONES

Respuestas de la Autoevaluación de la Unidad I

NÚMERO	RESPUESTA
1	VFVVFVVV
2	DADCB
3	Todos los frascos de cloro de esa marca, 100 botellas de cloro, contenido, 128 onzas, 126 onzas.
4	E
5	C

Respuestas de la Autoevaluación de la Unidad II

NÚMERO	RESPUESTA
1	· $\bar{x}$ =126,5 ; Me=126,2 ; Mo=126,2 · R=4,9 ; s=1,48 · La segunda muestra es más dispersa.
2	· P64=39,36 Dólares · P25=32,17 Dólares
3	La media es más representativa el día 20.
4	· Tipo 1: Asimétrico cola a la derecha. · Tipo 2: Asimétrico cola a la derecha.
5	· $\bar{x}$ =65 ; Me=65 ; Mo=65 · R=60 ; s2 = 130,43 ; s=11,42 ; CV=17,57% · As=0 ; K=0,2446

Respuestas de la Autoevaluación de la Unidad III

NÚMERO	RESPUESTA
1	· 2/6 = 0,333 · 2/4 = 0,5
2	2/4 = 0,5
3	4/8 = 0,5
4	(18/36)(10/35)(17/34) = 0,0714
5	0,180

Respuestas de la Autoevaluación de la Unidad IV

NÚMERO	RESPUESTA
1	· 0,0307 · 0,117 · 0,4418
2	0,44
3	0,4706
4	6,55%
5	139 soldados

Universidad
Continental

MANUALES AUTOFORMATIVOS INTERACTIVOS

Este manual autoformativo es el material didáctico más importante de la presente asignatura. Elaborado por el docente, orienta y facilita el auto aprendizaje de los contenidos y el desarrollo de las actividades propuestas en el sílabo.

Los demás recursos educativos del aula virtual complementan y se derivan del manual. Los contenidos multimedia ofrecidos utilizando videos, presentaciones, audios, clases interactivas, se corresponden a los contenidos del presente manual. La educación a distancia en entornos virtuales te permite estudiar desde el lugar donde te encuentres y a la hora que más te convenga. Basta conectarse al Internet, ingresar al campus virtual donde encontrarás todos tus servicios: aulas, videoclases, presentaciones animadas, biblioteca de recursos, muro y las tareas, siempre acompañado de tus docentes y amigos.

El modelo educativo de la Universidad Continental a distancia es innovador, interactivo e integral, conjugando el conocimiento, la investigación y la innovación. Su estructura, organización y funcionamiento están de acuerdo a los estándares internacionales. Es innovador, porque desarrolla las mejores prácticas del e-learning universitario global; interactivo, porque proporciona recursos para la comunicación y colaboración síncrona y asíncrona con docentes y estudiantes; e integral, pues articula contenidos, medios y recursos para el aprendizaje permanente y en espacios flexibles. Ahora podrás estar en la universidad en tiempo real sin ir a la universidad.