

# Notas de Estadística I

Carlos Martínez-Rodríguez  
Academia de Matemáticas  
Casa Libertad  
[carlos.martinez@uacm.edu.mx](mailto:carlos.martinez@uacm.edu.mx)

05 de febrero de 2019



Notas de un primer curso de Estadística para estudiantes de Modelación Matemática. Los contenidos a desarrollar son:

- ▶ Introducción
- ▶ Estimación de Parámetros poblacionales
- ▶ Comparación de dos medias y dos proporciones poblacionales
- ▶ Prueba no paramétrica
- ▶ Inferencia con tablas de contingencia
- ▶ Pruebas de Hipótesis

# Introducción

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

# Introducción

Como dijera Huntsberger: *La palabra estadística a menudo nos trae a la mente imágenes de números apilados en grandes arreglos y tablas, de volúmenes de cifras relativas a nacimientos, muertes, impuestos, poblaciones, ingresos, deudas, créditos y así sucesivamente.* Huntsberger tiene razón pues al instante de escuchar esta palabra estas son las imágenes que llegan a nuestra cabeza. La Estadística es mucho más que sólo números apilados y gráficas bonitas. Es una ciencia con tanta antiguedad como la escritura, y es por sí misma auxiliar de todas las demás ciencias. Los mercados, la medicina, la ingeniería, los gobiernos, etc. Se nombran entre los más destacados clientes de ésta. La ausencia de ésta conllevaría a un caos generalizado, dejando a los administradores y ejecutivos sin información vital a la hora de tomar decisiones en tiempos de incertidumbre.

# Introducción

La Estadística que conocemos hoy en día debe gran parte de su realización a los trabajos matemáticos de aquellos hombres que desarrollaron la teoría de las probabilidades, con la cual se adhirió a la Estadística a las ciencias formales.

## Definición

*La Estadística es la ciencia cuyo objetivo es reunir una información cuantitativa concerniente a individuos, grupos, series de hechos, etc. y deducir de ello gracias al análisis de estos datos unos significados precisos o unas previsiones para el futuro.*

La estadística, en general, es la ciencia que trata de la recopilación, organización presentación, análisis e interpretación de datos numéricos con el fin de realizar una toma de decisión más efectiva.

# Introducción

Otros autores la definen como la expresión cuantitativa del conocimiento dispuesta en forma adecuada para el escrutinio y análisis.

Los estudiantes confunden comúnmente los demás términos asociados con las Estadísticas, una confusión que es conveniente aclarar debido a que esta palabra tiene tres significados:

- ▶ la palabra estadística, en primer término se usa para referirse a la información estadística;
- ▶ también se utiliza para referirse al conjunto de técnicas y métodos que se utilizan para analizar la información estadística; y
- ▶ el término estadístico, en singular y en masculino, se refiere a una medida derivada de una muestra.

# Introducción

Los métodos estadísticos tradicionalmente se utilizan para propósitos descriptivos, para organizar y resumir datos numéricos. La estadística descriptiva, por ejemplo trata de la tabulación de datos, su presentación en forma gráfica o ilustrativa y el cálculo de medidas descriptivas.

Ahora bien, las técnicas estadísticas se aplican de manera amplia en mercadotecnia, contabilidad, control de calidad y en otras actividades; estudios de consumidores; análisis de resultados en deportes; administradores de instituciones; en la educación; organismos políticos; médicos; y por otras personas que intervienen en la toma de decisiones.

# Introducción: Historia

Es difícil conocer los orígenes de la Estadística. Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas. Su origen empieza posiblemente en la isla de Cerdeña, donde existen monumentos prehistóricos pertenecientes a los Nuragas, las primeros habitantes de la isla; estos monumentos constan de bloques de basalto superpuestos sin mortero y en cuyas paredes de encontraban grabados toscos signos que han sido interpretados con mucha verosimilitud como muescas que servían para llevar la cuenta del ganado y la caza.

# Introducción: Historia

Los babilonios usaban ya pequeñas tablillas de arcilla para recopilar datos en tablas sobre la producción agrícola y los géneros vendidos o cambiados mediante trueque. Por otra parte, otros vestigios pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo, prolíficos datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, dicho registro de riqueza y población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo de las tierras con el objeto de verificar un nuevo reparto.

# Introducción: Historia

En el antiguo Israel la Biblia da referencias, en el libro de los Números, de los datos estadísticos obtenidos en dos recuentos de la población hebrea. El rey David por otra parte, ordenó a Joab, general del ejército hacer un censo de Israel con la finalidad de conocer el número de la población.

También los chinos efectuaron censos hace más de cuarenta siglos. Los griegos efectuaron censos periódicamente con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). La investigación histórica revela que se realizaron 69 censos para calcular los impuestos, determinar los derechos de voto y ponderar la potencia guerrera.

# Introducción: Historia

Pero fueron los romanos, maestros de la organización política, quienes mejor supieron emplear los recursos de la estadística. Cada cinco años realizaban un censo de la población y sus funcionarios públicos tenían la obligación de anotar nacimientos, defunciones y matrimonios, sin olvidar los recuentos periódicos del ganado y de las riquezas contenidas en las tierras conquistadas. Para el nacimiento de Cristo sucedía uno de estos empadronamientos de la población bajo la autoridad del imperio.

# Introducción: Historia

Durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones Estadísticas, con la notable excepción de las relaciones de tierras pertenecientes a la Iglesia, compiladas por Pipino el Breve en el 758 y por Carlomagno en el 762 DC. Durante el siglo IX se realizaron en Francia algunos censos parciales de siervos. En Inglaterra, Guillermo el Conquistador recopiló el Domesday Book o libro del Gran Catastro para el año 1086, un documento de la propiedad, extensión y valor de las tierras de Inglaterra. Esa obra fue el primer compendio estadístico de Inglaterra. Aunque Carlomagno, en Francia; y Guillermo el Conquistador, en Inglaterra, trataron de revivir la técnica romana, los métodos estadísticos permanecieron casi olvidados durante la Edad Media.

# Introducción: Historia

Durante los siglos XV, XVI, y XVII, hombres como Leonardo de Vinci, Nicolás Copérnico, Galileo, Neper, William Harvey, Sir Francis Bacon y René Descartes, hicieron grandes operaciones al método científico, de tal forma que cuando se crearon los Estados Nacionales y surgió como fuerza el comercio internacional existía ya un método capaz de aplicarse a los datos económicos. Para el año 1532 empezaron a registrarse en Inglaterra las defunciones debido al temor que Enrique VII tenía por la peste. Más o menos por la misma época, en Francia la ley exigió a los clérigos registrar los bautismos, fallecimientos y matrimonios. Durante un brote de peste que apareció a fines de la década de 1500, el gobierno inglés comenzó a publicar estadística semanales de los decesos. Esta costumbre continuó muchos años, y en 1632 estos Bills of Mortality (Cuentas de Mortalidad) contenían los nacimientos y fallecimientos por sexo.

# Introducción: Historia

En 1662, el capitán John Graunt usó documentos que abarcaban treinta años y efectuó predicciones sobre el número de personas que morirían de varias enfermedades y sobre las proporciones de nacimientos de varones y mujeres que cabría esperar. El trabajo de Graunt, condensado en su obra *Natural and Political Observations...Made upon the Bills of Mortality*, fue un esfuerzo innovador en el análisis estadístico.

Por el año 1540 el alemán Sebastián Muster realizó una compilación estadística de los recursos nacionales, comprensiva de datos sobre organización política, instrucciones sociales, comercio y poderío militar. Durante el siglo XVII aportó indicaciones más concretas de métodos de observación y análisis cuantitativo y amplió los campos de la inferencia y la teoría Estadística.

# Introducción: Historia

Los eruditos del siglo XVII demostraron especial interés por la Estadística Demográfica como resultado de la especulación sobre si la población aumentaba, decrecía o permanecía estática. En los tiempos modernos tales métodos fueron resucitados por algunos reyes que necesitaban conocer las riquezas monetarias y el potencial humano de sus respectivos países. El primer empleo de los datos estadísticos para fines ajenos a la política tuvo lugar en 1691 y estuvo a cargo de Gaspar Neumann, un profesor alemán que vivía en Breslau. Este investigador se propuso destruir la antigua creencia popular de que en los años terminados en siete moría más gente que en los restantes, y para lograrlo hurgó pacientemente en los archivos parroquiales de la ciudad. Después de revisar miles de partidas de defunción pudo demostrar que en tales años no fallecían más personas que en los demás.

# Introducción: Historia

Los procedimientos de Neumann fueron conocidos por el astrónomo inglés Halley, descubridor del cometa que lleva su nombre, quien los aplicó al estudio de la vida humana. Sus cálculos sirvieron de base para las tablas de mortalidad que hoy utilizan todas las compañías de seguros. Durante el siglo XVII y principios del XVIII, matemáticos como Bernoulli, Francis Maseres, Lagrange y Laplace desarrollaron la teoría de probabilidades. No obstante durante cierto tiempo, la teoría de las probabilidades limitó su aplicación a los juegos de azar y hasta el siglo XVIII no comenzó a aplicarse a los grandes problemas científicos.

# Introducción: Historia

Godofredo Achenwall, profesor de la Universidad de Gotinga, acuñó en 1760 la palabra estadística, que extrajo del término italiano statista (estadista). Creía, y con sobrada razón, que los datos de la nueva ciencia serían el aliado más eficaz del gobernante consciente. La raíz remota de la palabra se halla, por otra parte, en el término latino status, que significa estado o situación; Esta etimología aumenta el valor intrínseco de la palabra, por cuanto la estadística revela el sentido cuantitativo de las más variadas situaciones. Jacques Quételet es quien aplica las Estadísticas a las ciencias sociales. Este interpretó la teoría de la probabilidad para su uso en las ciencias sociales y resolver la aplicación del principio de promedios y de la variabilidad a los fenómenos sociales.

# Introducción: Historia

Quételect fue el primero en realizar la aplicación práctica de todo el método Estadístico, entonces conocido, a las diversas ramas de la ciencia.

Entretanto, en el período del 1800 al 1820 se desarrollaron dos conceptos matemáticos fundamentales para la teoría Estadística; la teoría de los errores de observación, aportada por Laplace y Gauss; y la teoría de los mínimos cuadrados desarrollada por Laplace, Gauss y Legendre. A finales del siglo XIX, Sir Francis Gaston ideó el método conocido por Correlación, que tenía por objeto medir la influencia relativa de los factores sobre las variables.

# Introducción: Historia

De aquí partió el desarrollo del coeficiente de correlación creado por Karl Pearson y otros cultivadores de la ciencia biométrica como J. Pease Norton, R. H. Hooker y G. Udny Yule, que efectuaron amplios estudios sobre la medida de las relaciones.

Los progresos más recientes en el campo de la Estadística se refieren al ulterior desarrollo del cálculo de probabilidades, particularmente en la rama denominada indeterminismo o relatividad, se ha demostrado que el determinismo fue reconocido en la Física como resultado de las investigaciones atómicas y que este principio se juzga aplicable tanto a las ciencias sociales como a las físicas.

# Introducción: Historia

La historia de la estadística está resumida en tres grandes etapas o fases.

**Fase 1: Los Censos:** Desde el momento en que se constituye una autoridad política, la idea de inventariar de una forma más o menos regular la población y las riquezas existentes en el territorio está ligada a la conciencia de soberanía y a los primeros esfuerzos administrativos.

## Fase 2: De la Descripción de los Conjuntos a la Aritmética

**Política:** Las ideas mercantilistas extrañan una intensificación de este tipo de investigación. Colbert multiplica las encuestas sobre artículos manufacturados, el comercio y la población: los intendentes del Reino envían a París sus memorias. Vauban, más conocido por sus fortificaciones o su Dime Royale, que es la primera propuesta de un impuesto sobre los ingresos, se señala como el verdadero precursor de los sondeos. Más tarde, Bufón se preocupa de esos problemas antes de dedicarse a la historia natural. La escuela inglesa proporciona un nuevo progreso al superar la fase puramente descriptiva.

# Introducción

**Fase 2:** Sus tres principales representantes son Graunt, Petty y Halley. El penúltimo es autor de la famosa Aritmética Política. Chaptal, ministro del interior francés, publica en 1801 el primer censo general de población, desarrolla los estudios industriales, de las producciones y los cambios, haciéndose sistemáticos durante las dos terceras partes del siglo XIX.

**Fase 3:** **Estadística y Cálculo de Probabilidades:** El cálculo de probabilidades se incorpora rápidamente como un instrumento de análisis extremadamente poderoso para el estudio de los fenómenos económicos y sociales y en general para el estudio de fenómenos cuyas causas son demasiados complejas para conocerlos totalmente y hacer posible su análisis.

# Introducción: Estadística Descriptiva

La Estadística para su mejor estudio se ha dividido en dos grandes ramas: **la Estadística Descriptiva y la Estadística Inferencial.**

- ▶ **Descriptiva:** consiste sobre todo en la presentación de datos en forma de tablas y gráficas. Esta comprende cualquier actividad relacionada con los datos y está diseñada para resumir o describir los mismos sin factores pertinentes adicionales; esto es, sin intentar inferir nada que vaya más allá de los datos, como tales.

# Introducción: Estadística Inferencial

- ▶ **Inferencial:** se deriva de muestras, de observaciones hechas sólo acerca de una parte de un conjunto numeroso de elementos y esto implica que su análisis requiere de generalizaciones que van más allá de los datos. Como consecuencia, la característica más importante del reciente crecimiento de la estadística ha sido un cambio en el énfasis de los métodos que describen a métodos que sirven para hacer generalizaciones. La Estadística Inferencial investiga o analiza una población partiendo de una muestra tomada.

# Introducción: Estadística Inferencial

Los métodos básicos de la estadística inferencial son la estimación y el contrastede hipótesis, que juegan un papel fundamental en la investigación. Por tanto, algunos de los objetivos que se persiguen son:

- ▶ Calcular los parámetros de la distribución de medias o proporciones muestrales de tamaño  $n$ , extraídas de una población de media y varianza conocidas.
- ▶ Estimar la media o la proporción de una población a partir de la media o proporción muestral.
- ▶ Utilizar distintos tamaños muestrales para controlar la confianza y el error admitido.
- ▶ Contrastar los resultados obtenidos a partir de muestras.
- ▶ Visualizar gráficamente, mediante las respectivas curvas normales, las estimaciones realizadas.

# Introducción: Estadística Inferencial

En definitiva, la idea es, a partir de una población se extrae una muestra por algunos de los métodos existentes, con la que se generan datos numéricos que se van a utilizar para generar estadísticos con los que realizar estimaciones o contrastes poblacionales.

Existen dos formas de estimar parámetros: la *estimación puntual* y la *estimación por intervalo de confianza*. En la primera se busca, con base en los datos muestrales, un único valor estimado para el parámetro. Para la segunda, se determina un intervalo dentro del cual se encuentra el valor del parámetro, con una probabilidad determinada.

# Introducción: Estimación de Parámetros

Si el objetivo del tratamiento estadístico inferencial, es efectuar generalizaciones acerca de la estructura, composición o comportamiento de las poblaciones no observadas, a partir de una parte de la población, será necesario que la parcela de población examinada sea representativa del total.

Por ello, la selección de la muestra requiere unos requisitos que lo garanticen, debe ser representativa y aleatoria.

Además, la cantidad de elementos que integran la muestra (el tamaño de la muestra) depende de múltiples factores, como el dinero y el tiempo disponibles para el estudio, la importancia del tema analizado, la confiabilidad que se espera de los resultados, las características propias del fenómeno analizado, etcétera.

# Introducción: Método Estadístico

Así, a partir de la muestra seleccionada se realizan algunos cálculos y se estima el valor de los parámetros de la población tales como la media, la varianza, la desviación estándar, o la forma de la distribución, etc.

# Introducción: Método Estadístico

El conjunto de los métodos que se utilizan para medir las características de la información, para resumir los valores individuales, y para analizar los datos a fin de extraerles el máximo de información, es lo que se llama *métodos estadísticos*. Los métodos de análisis para la información cuantitativa se pueden dividir en los siguientes seis pasos:

1. Definición del problema.
2. Recopilación de la información existente.
3. Obtención de información original.
4. Clasificación.
5. Presentación.
6. Análisis.

# Introducción: Método Estadístico

El centro de gravedad de la metodología estadística se empieza a desplazar técnicas de computación intensiva aplicadas a grandes masas de datos, y se empieza a considerar el método estadístico como un proceso iterativo de búsqueda del modelo ideal.

Las aplicaciones en este periodo de la Estadística a la Economía conducen a una disciplina con contenido propio: la Econometría. La investigación estadística en problemas militares durante la segunda guerra mundial y los nuevos métodos de programación matemática, dan lugar a la Investigación Operativa

# Introducción: Errores Estadísticos

Al momento de recopilar los datos que serán procesados se es susceptible de cometer errores así como durante los cómputos de los mismos. No obstante, hay otros errores que no tienen nada que ver con la digitación y que no son tan fácilmente identificables.

Algunos de éstos errores son:

- ▶ **Sesgo:** Es imposible ser completamente objetivo o no tener ideas preconcebidas antes de comenzar a estudiar un problema, y existen muchas maneras en que una perspectiva o estado mental pueda influir en la recopilación y en el análisis de la información. En estos casos se dice que hay un sesgo cuando el individuo da mayor peso a los datos que apoyan su opinión que a aquellos que la contradicen. Un caso extremo de sesgo sería la situación donde primero se toma una decisión y después se utiliza el análisis estadístico para justificar la decisión ya tomada.

# Introducción: Método Estadístico

- ▶ **Datos No Comparables:** el establecer comparaciones es una de las partes más importantes del análisis estadístico, pero es extremadamente importante que tales comparaciones se hagan entre datos que sean comparables.
- ▶ **Proyección descuidada de tendencias:** la proyección simplista de tendencias pasadas hacia el futuro es uno de los errores que más ha desacreditado el uso del análisis estadístico.
- ▶ **Muestreo Incorrecto:** en la mayoría de los estudios sucede que el volumen de información disponible es tan inmenso que se hace necesario estudiar muestras, para derivar conclusiones acerca de la población a la que pertenece la muestra. Si la muestra se selecciona correctamente, tendrá básicamente las mismas propiedades que la población de la cual fue extraída; pero si el muestreo se realiza incorrectamente, entonces puede suceder que los resultados no signifiquen nada

# Introducción: Método Estadístico

Un conjunto de procedimientos para reunir, clasificar, codificar, procesar, analizar y resumir información numérica adquirida sistemáticamente (Ritchey, 2002). Permite hacer inferencias a partir de una muestra para extrapolarlas a una población. Aunque normalmente se asocia a muchos cálculos y operaciones aritméticas, y aunque las matemáticas están involucradas, en su mayor parte sus fundamentos y uso apropiado pueden dominarse sin hacer referencia a habilidades matemáticas avanzadas.

De hecho se trata de una forma de ver la realidad basada en el análisis cuidadoso de los hechos (Ritchey, 2002). Es necesaria sin embargo la sistematización para reducir el efecto que las emociones y las experiencias individuales puedan tener al interpretar esa realidad.

## Introducción: Refuerzo

De esta manera la estadística se relaciona con el método científico complementándolo como herramienta de análisis y, aunque la investigación científica no requiere necesariamente de la estadística, ésta valida muchos de los resultados cuantitativos derivados de la investigación.

La obtención del conocimiento debe hacerse de manera sistemática por lo que deben planearse todos los pasos que llevan desde el planteamiento de un problema, pasando por la elaboración de hipótesis y la manera en que van a ser probadas; la selección de sujetos (muestreo), los escenarios, los instrumentos que se utilizarán para obtener los datos, definir el procedimiento que se seguirá para esto último, los controles que se deben hacer para asegurar que las intervenciones son las causas más probables de los cambios esperados (diseño);

# Introducción: Refuerzo

hasta la elección del plan de análisis idóneo para el tipo de datos que se están obteniendo, es aquí donde la estadística entra en el estudio, aunque pueden existir otras herramientas de análisis si se está haciendo una investigación de corte cualitativo.

Una buena planeación permitirá que los resultados puedan ser reproducidos, mediante la comprobación empírica, por cualquier investigador interesado en refutar o comprobar las conclusiones que se hagan del estudio. De esta manera también se logrará la predicción de los fenómenos que se están estudiando, ayudando a conocer y prevenir los problemas sociales e individuales que forman parte del objeto de estudio de la psicología.

# Introducción: Refuerzo

El tratamiento de los datos de la investigación científica tiene varias etapas:

- ▶ En la etapa de recolección de datos del método científico, se define a la población de interés y se selecciona una muestra o conjunto de personas representativas de la misma, se realizan experimentos o se emplean instrumentos ya existentes o de nueva creación, para medir los atributos de interés necesarios para responder a las preguntas de investigación. Durante lo que es llamado trabajo de campo se obtienen los datos en crudo, es decir las respuestas directas de los sujetos uno por uno, se codifican (se les asignan valores a las respuestas), se capturan y se verifican para ser utilizados en las siguientes etapas.

# Introducción: Refuerzo

- ▶ En la etapa de recuento, se organizan y ordenan los datos obtenidos de la muestra. Esta será descrita en la siguiente etapa utilizando la estadística descriptiva, todas las investigaciones utilizan estadística descriptiva, para conocer de manera organizada y resumida las características de la muestra.
- ▶ En la etapa de análisis se utilizan las pruebas estadísticas (estadística inferencial) y en la interpretación se acepta o rechaza la hipótesis nula.

# Introducción: Refuerzo

- ▶ En investigación, el fenómeno en estudio puede ser cualitativo que implicaría comprenderlo y explicarlo, o cuantitativo para compararlo y hacer inferencias. Se puede decir que si se hace análisis se usan métodos cuantitativos y si se hace descripción se usan métodos cualitativos. Medición Para poder emplear el método estadístico en un estudio es necesario medir las variables.
  - ▶ Medir: es asignar valores a las propiedades de los objetos bajo ciertas reglas, esas reglas son los niveles de medición
  - ▶ Cuantificar: es asignar valores a algo tomando un patrón de referencia. Por ejemplo, cuantificar es ver cuántos hombres y cuántas mujeres hay.

# Introducción: Refuerzo

- ▶ Variable: es una característica o propiedad que asume diferentes valores dentro de una población de interés y cuya variación es susceptible de medirse. Las variables pueden clasificarse de acuerdo al tipo de valores que puede tomar como:
  - ▶ Discretas o categóricas.- en las que los valores se relacionan a nombres, etiquetas o categorías, no existe un significado numérico directo
  - ▶ Continuas.- los valores tienen un correlato numérico directo, son continuos y susceptibles de fraccionarse y de poder utilizarse en operaciones aritméticas De acuerdo a la cantidad de valores
  - ▶ Dicotómica.- sólo tienen dos valores posibles, la característica está ausente o presente
  - ▶ Policotómica.- pueden tomar tres valores o más, pueden tomarse matices diferentes, en grados, jerarquías o magnitudes continuas.

# Introducción: Refuerzo

- ▶ En cuanto a una clasificación estadística
  - ▶ Aleatoria.- Aquella en la cual desconocemos el valor porque fluctúa de acuerdo a un evento debido al azar
  - ▶ Determinística.- Aquella variable de la que se conoce el valor
  - ▶ Independiente.- aquellas variables que son manipuladas por el investigador. Define los grupos
  - ▶ Dependiente.- son mediciones que ocurren durante el experimento o tratamiento (resultado de la independiente), es la que se mide y compara entre los grupos

# Introducción: Refuerzo

## Niveles de Medición

- ▶ Nominal Las propiedades de la medición nominal son:
  - ▶ Exhaustiva: implica a todas las opciones
  - ▶ A los sujetos se les asignan categorías, por lo que son mutuamente excluyentes. Es decir, la variable está presente o no; tiene o no una característica
- ▶ Ordinal Las propiedades de la medición ordinal son:
  - ▶ El nivel ordinal posee transitividad, por lo que se tiene la capacidad de identificar que es mejor o mayor que otra, en ese sentido se pueden establecer jerarquías
  - ▶ Las distancias entre un valor y otro no son iguales.

# Introducción: Refuerzo

## ▶ Intervalo

- ▶ El nivel de medición intervalar requiere distancias iguales entre cada valor. Por lo general utiliza datos cuantitativos. Por ejemplo: temperatura, atributos psicológicos (CI, nivel de autoestima, pruebas de conocimientos, etc.)
- ▶ Las unidades de calificación son equivalentes en todos los puntos de la escala. Una escala de intervalos implica: clasificación, magnitud y unidades de tamaños iguales (Brown, 2000).
- ▶ Se pueden hacer operaciones aritméticas
- ▶ Cuando se le pide al sujeto que califique una situación del 0 al 10 puede tomarse como un nivel de medición de intervalo, siempre y cuando se incluya el 0.

# Introducción: Refuerzo

## ► Razón

- La escala empieza a partir del 0 absoluto, por lo tanto incluye sólo los números por su valor en sí, por lo que no pueden existir los números con signo negativo. Por ejemplo: Peso corporal en kg., edad en años, estatura en cm.
- Convencionalmente los datos que son de nivel absoluto o de razón son manejados como los datos intervalares.

Las variables intervalares y de razón se pueden incluir en operaciones aritméticas y para efectos de análisis estadísticos se consideran con las mismas propiedades y se clasifican en un rubro general llamado nivel escalar. Las variables ordinales, por otro lado, aparentemente no se pueden sumar, pero cuando se tiene una escala con reactivos calificados en un nivel ordinal, al ser sumados el resultado de esta suma puede ser considerado como intervalar, ya que matemáticamente se comporta como tal.

# Introducción: Refuerzo

- ▶ **Variable:** Consideraciones que una variable son una característica o fenómeno que puede tomar distintos valores.
- ▶ **Dato:** Mediciones o cualidades que han sido recopiladas como resultado de observaciones.
- ▶ **Población:** Se considera el área de la cual son extraídos los datos. Es decir, es el conjunto de elementos o individuos que poseen una característica común y medible acerca de lo cual se desea información. Es también llamado Universo.
- ▶ **Muestra:** Es un subconjunto de la población, seleccionado de acuerdo a una regla o algún plan de muestreo.
- ▶ **Censo:** Recopilación de todos los datos (de interés para la investigación) de la población.
- ▶ **Estadística:** Es una función o fórmula que depende de los datos de la muestra (es variable).

# Introducción: Refuerzo

- ▶ **Parámetro:** Característica medible de la población. Es un resumen numérico de alguna variable observada de la población. Los parámetros normales que se estudian son: *La media poblacional, la media poblacional, Proporción.*
- ▶ **Estimador:** Un estimador de un parámetro , es un estadístico que se emplea para conocer el parámetro desconocido.
- ▶ **Estadístico:** Es una función de los valores de la muestra. Es una variable aleatoria, cuyos valores dependen de la muestra seleccionada. Su distribución de probabilidad, se conoce como *Distribución muestral del estadístico.*

# Introducción: Refuerzo

- ▶ **Estimación:** Este término indica que a partir de lo observado en una muestra (un resumen estadístico con las medidas que conocemos de Descriptiva) se extrapola o generaliza dicho resultado muestral a la población total, de modo que lo estimado es el valor generalizado a la población. Consiste en la búsqueda del valor de los parámetros poblacionales objeto de estudio. Puede ser puntual o por intervalo de confianza:
  - ▶ *Puntual:* cuando buscamos un valor concreto. Un estimador de un parámetro poblacional es una función de los datos muestrales. En pocas palabras, es una fórmula que depende de los valores obtenidos de una muestra, para realizar estimaciones. Lo que se pretende obtener es el valor exacto de un parámetro.

# Introducción: Refuerzo

Las propiedades deseables de un estimador son los siguientes:

- ▶ Insesgado: Diremos que un estimador de un parámetro es insesgado si su esperanza coincide con el verdadero valor del parámetro. En el caso de que no coincidan, diremos que el estimador es sesgado.
- ▶ Eficiencia: Dados dos estimadores para un mismo parámetro , se dice que uno es má eficiente que el otro si tiene menor varianza.
- ▶ Suficiencia: Se dice que un estimador de un parámetro es suficiente cuando para su cálculo utiliza toda la información de la muestra.
- ▶ Consistencia: Decimos que un estimador de un parámetro es consistente si la distribución del estimador tiende a concentrarse en un cierto punto cuando el tamaño de la muestra tiende a infinito.

## Introducción: Refuerzo

Demostrar que un cierto estimador cumple estas propiedades puede ser complicado en determinadas ocasiones. Existen varios métodos que nos van a permitir obtener los estimadores puntuales. Los más importantes son:

- ▶ Método de Momentos: se basa en que los momentos poblacionales y se estiman mediante los momentos muestrales. Suelen dar estimadores consistentes.
- ▶ Método de Mínimos Cuadrados: consiste en obtener un estimador que hace mínima una determinada función.
- ▶ Método de Máxima Verosimilitud: consiste en tomar como parámetro poblacional el valor de la muestra que sea más probable, es decir, que tenga mayor probabilidad. Se suelen obtener estimadores consistentes y eficientes. Es el más utilizado.

# Introducción: Refuerzo

- ▶ *Intervalo de confianza:* cuando determinamos un intervalo, dentro del cual se supone que va a estar el valor del parámetro que se busca con una cierta probabilidad. El intervalo de confianza está determinado por dos valores dentro de los cuales afirmamos que está el verdadero parámetro con cierta probabilidad. Son unos límites o margen de variabilidad que damos al valor estimado, para poder afirmar, bajo un criterio de probabilidad, que el verdadero valor no los rebasará. Este intervalo contiene al parámetro estimado con una determinada certeza o nivel de confianza.

## Introducción: Refuerzo

En la estimación por intervalos se usan los siguientes conceptos:

- ▶ Variabilidad del parámetro: Si no se conoce, puede obtenerse una aproximación en los datos o en un estudio piloto. También hay métodos para calcular el tamaño de la muestra que prescinden de este aspecto. Habitualmente se usa como medida de esta variabilidad la desviación típica poblacional.
- ▶ Error de la estimación: Es una medida de su precisión que se corresponde con la amplitud del intervalo de confianza. Cuanta más precisión se desee en la estimación de un parámetro, más estrecho deberá ser el intervalo de confianza y, por tanto, menor el error, y más sujetos deberán incluirse en la muestra estudiada.
- ▶ Nivel de confianza: Es la probabilidad de que el verdadero valor del parámetro estimado en la población se sitúe en el intervalo de confianza obtenido. El nivel de confianza se denota por  $1 - \alpha$

# Introducción: Refuerzo

- ▶  $p$ -value : También llamado nivel de significación. Es la probabilidad (en tanto por uno) de fallar en nuestra estimación, esto es, la diferencia entre la certeza (1) y el nivel de confianza  $1 - \alpha$ .
- ▶ Valor crítico: Se representa por  $Z_{\alpha/2}$ . Es el valor de la abscisa en una determinada distribución que deja a su derecha un área igual a  $1/2$ , siendo  $1 - \alpha$  el nivel de confianza. Normalmente los valores críticos están tabulados o pueden calcularse en función de la distribución de la población.

# Introducción: Refuerzo

Para un tamaño fijo de la muestra, los conceptos de error y nivel de confianza van relacionados. Si admitimos un error mayor, esto es, aumentamos el tamaño del intervalo de confianza, tenemos también una mayor probabilidad de éxito en nuestra estimación, es decir, un mayor nivel de confianza. Por tanto, un aspecto que debe de tenerse en cuenta es el tamaño muestral, ya que para disminuir el error que se comete habrá que aumentar el tamaño muestral. Esto se resolverá, para un intervalo de confianza cualquiera, despejando el tamaño de la muestra en cualquiera de las formulas de los intervalos de confianza que veremos a continuación, a partir del error máximo permitido. Los intervalos de confianza pueden ser unilaterales o bilaterales:

# Introducción: Refuerzo

- ▶ **Contraste de Hipótesis:** Consiste en determinar si es aceptable, partiendo de datos muestrales, que la característica o el parámetro poblacional estudiado tome un determinado valor o esté dentro de unos determinados valores.
- ▶ **Nivel de Confianza:** Indica la proporción de veces que acertaríamos al afirmar que el parámetro está dentro del intervalo al seleccionar muchas muestras.

## Introducción: Refuerzo

Una muestra es representativa en la medida que es imagen de la población.

En general, podemos decir que el tamaño de una muestra dependerá principalmente de: *Nivel de precisión deseado, Recursos disponibles, Tiempo involucrado en la investigación..* Además el plan de muestreo debe considerar *La población, Parámetros a medir.*

Existe una gran cantidad de tipos de muestreo. En la práctica los más utilizados son los siguientes:

- ▶ **MUESTREO ALEATORIO SIMPLE:** Es un método de selección de  $n$  unidades extraídas de  $N$ , de tal manera que cada una de las posibles muestras tiene la misma probabilidad de ser escogida. (En la práctica, se enumeran las unidades de 1 a  $N$ , y a continuación se seleccionan  $n$  números aleatorios entre 1 y  $N$ , ya sea de tablas o de alguna urna con fichas numeradas).

# Introducción: Refuerzo

- ▶ **MUESTREO ESTRATIFICADO ALEATORIO:** Se usa cuando la población está agrupada en pocos estratos, cada uno de ellos son muchas entidades. Este muestreo consiste en sacar una muestra aleatoria simple de cada uno de los estratos. (Generalmente, de tamaño proporcional al estrato).
- ▶ **MUESTREO SISTEMÁTICO :** Se utiliza cuando las unidades de la población están de alguna manera totalmente ordenadas. Para seleccionar una muestra de  $n$  unidades, se divide la población en  $n$  subpoblaciones de tamaño  $K = N/n$  y se toma al azar una unidad de la  $K$  primeras y de ahí en adelante cada  $K$ -ésima unidad.
- ▶ **MUESTREO POR CONGLOMERADO:** Se emplea cuando la población está dividida en grupos o conglomerados pequeños. Consiste en obtener una muestra aleatoria simple de conglomerados y luego CENSAR cada uno de éstos.

# Introducción: Refuerzo

- ▶ **MUESTREO EN DOS ETAPAS (Bietápico):** En este caso la muestra se toma en dos pasos:
  - ▶ Seleccionar una muestra de unidades primarias, y Seleccionar una muestra de elementos a partir de cada unidad primaria escogida.
  - ▶ Observación: En la realidad es posible encontrarse con situaciones en las cuales no es posible aplicar libremente un tipo de muestreo, incluso estaremos obligados a mezclarlas en ocasiones.

## Introducción: Refuerzo

Las variables se pueden clasificar en dos grandes grupos.

- ▶ **Variables categóricas:** Son aquellas que pueden ser representadas a través de símbolos, letras, palabras, etc. Los valores que toman se denominan categorías, y los elementos que pertenecen a estas categorías, se consideran idénticos respecto a la característica que se está midiendo. Las variables categóricas se dividen en dos tipos: Ordinal y Nominal.
    - ▶ **Las Ordinales**, son aquellas en las que las categorías tienen un orden implícito. Admiten grados de calidad, es decir, existe una relación total entre las categorías.
    - ▶ **Las nominales**, son aquellas donde no existe una relación de orden.

# Introducción: Refuerzo

- ▶ **Variables Numéricas:** Son aquellas que pueden tomar valores numéricos exclusivamente (mediciones). Se dividen en dos tipos: Discretas y continuas.
  - ▶ **Discretas:** son aquellas que toman sus valores en un conjunto finito o infinito numerable.
  - ▶ **Continuas:** Son aquellas que toman sus valores en un subconjunto de los números reales, es decir en un intervalo. En general para las variables continuas el hombre ha debido inventar una medida para poder establecer una medición de ellas.

## Introducción: Refuerzo

El propósito de esta sección es solamente indicar los malos usos comunes de datos estadísticos, sin incluir el uso de métodos estadísticos complicados. Un estudiante debería estar alerta en relación con estos malos usos y debería hacer un gran esfuerzo para evitarlos a fin de ser un verdadero estadístico.

**Datos estadísticos inadecuados** Los datos estadísticos son usados como la materia prima para un estudio estadístico. Cuando los datos son inadecuados, la conclusión extraída del estudio de los datos se vuelve obviamente inválida. Por ejemplo, supongamos que deseamos encontrar el ingreso familiar típico del año pasado en la ciudad Y de 50,000 familias y tenemos una muestra consistente del ingreso de solamente tres familias: 1 millón, 2 millones y no ingreso. Si sumamos el ingreso de las tres familias y dividimos el total por 3, obtenemos un promedio de 1 millón.

## Introducción: Refuerzo

Entonces, extraemos una conclusión basada en la muestra de que el ingreso familiar promedio durante el año pasado en la ciudad fue de 1 millón.

Es obvio que la conclusión es falsa, puesto que las cifras son extremas y el tamaño de la muestra es demasiado pequeño; por lo tanto la muestra no es representativa.

Hay muchas otras clases de datos inadecuados. Por ejemplo, algunos datos son respuestas inexactas de una encuesta, porque las preguntas usadas en la misma son vagas o engañosas, algunos datos son toscas estimaciones porque no hay disponibles datos exactos o es demasiado costosa su obtención, y algunos datos son irrelevantes en un problema dado, porque el estudio estadístico no está bien planeado.

# Introducción: Refuerzo

Sesgo significa que un usuario dé los datos perjudicialmente de más énfasis a los hechos, los cuales son empleados para mantener su predeterminada posición u opinión.

Los estadísticos son frecuentemente degradados por lemas tales como :Hay tres clases de mentiras: mentiras, mentiras reprobables y estadística, y Las cifras no mienten, pero los mentirosos piensan.

# Introducción: Sesgo

Hay dos clases de sesgos: conscientes e inconscientes. Ambos son comunes en el análisis estadístico. Hay numerosos ejemplos de sesgos conscientes.

Un anunciante frecuentemente usa la estadística para probar que su producto es muy superior al producto de su competidor.

Un político prefiere usar la estadística para sostener su punto de vista.

Gerentes y líderes de trabajadores pueden simultáneamente situar sus respectivas cifras estadísticas sobre la misma tabla de trato para mostrar que sus rechazos o peticiones son justificadas.

# Introducción: Refuerzo

Es casi imposible que un sesgo inconsciente esté completamente ausente en un trabajo estadístico.

En lo que respecta al ser humano, es difícil obtener una actitud completamente objetiva al abordar un problema, aun cuando un científico debería tener una mente abierta.

Un estadístico debería estar enterado del hecho de que su interpretación de los resultados del análisis estadístico está influenciado por su propia experiencia, conocimiento y antecedentes con relación al problema dado.

# Introducción: Refuerzo

Es muy frecuente que un análisis estadístico contemple supuestos. Un investigador debe ser muy cuidadoso en este hecho, para evitar que éstos sean falsos. Los supuestos falsos pueden ser originados por:

- ▶ Quien usa los datos
- ▶ Quien está tratando de confundir (con intencionalidad)
- ▶ Ignorancia
- ▶ Descuido.