

Statistics and occupational health: basic statistics concepts

Estadística y Salud en el Trabajo: conceptos básicos de estadística

Juan Luis Soto Espinosa¹.

(1) FES Zaragoza, UNAM

Edificio de Posgrado, planta baja, cubículo 17, FES Zaragoza, Campus II. Av. Batalla de 5 de mayo esq. Fuerte de Loreto Col Ejército de Oriente. C.P. 09230 Iztapalapa, CDMX

Correo electrónico de contacto: soej@unam.mx

Fecha de envío: 15/06/2019

Fecha de aprobación: 02/09/2019

Abstract

Statistics is the discipline that is responsible for the collection, organization, analysis and presentation of data from the study of natural, social phenomena or productive, administrative, health processes among many others. This discipline aims to understand, with a certain degree of confidence, the behavior of the variables under study and to create models that allow predicting future conditions based on the behavior of the data collected. In this first installment, basic concepts of the discipline are presented, essential for the reader interested in the area to enter this field and deepen the use of techniques and tools for the analysis of data that will be presented in future deliveries.

Keywords: statistics, variable, population, sample, types of statistics

Resumen

La estadística es la disciplina que se encarga del acopio, organización, análisis y presentación de datos provenientes del estudio de fenómenos naturales, sociales o procesos productivos, administrativos, de salud entre muchos otros. Esta disciplina pretende entender, con cierto grado de confianza, el comportamiento de las variables motivo de estudio y de crear modelos que permitan predecir condiciones futuras a partir del comportamiento de los datos colectados. En esta primera entrega, se presentan conceptos básicos de la disciplina, indispensables para que el lector interesado en el área incursione en este campo y profundice en el uso de técnicas y herramientas para el análisis de datos que se presentarán en entregas futuras.

Palabras clave: Estadística, variables, población, muestra, tipos de estadística.

En esta edición iniciamos una serie dedicada a abordar aspectos básicos de estadística enfocados al quehacer cotidiano de las Ciencias de la Salud, específicamente la Salud en el Trabajo.

Empecemos definiendo el ámbito de la Estadística como disciplina. Antes que nada, es necesario notar que ESTADÍSTICA viene de ESTADO, ya que

desde que los seres humanos se integran en sociedades organizadas, es función de los gobiernos de los estados llevar el registro de la población (nacimientos, defunciones, matrimonios, individuos que se integran como nuevos ciudadanos, etc.) así como de los productos que se generan y los bienes con que se cuenta (cosechas, ganado, pesca, producción minera, bosques, tierras, impuestos,

Documento educativo

entre otros). Al irse haciendo cada vez más grandes y complejas estas sociedades, resulta indispensable contar con mecanismos de conteo, acopio y organización de datos que permitan conocer las cantidades existentes, estimar la producción y, sobre todo, contar con herramientas que faciliten la administración y la previsión de contingencias futuras.

A lo largo de la historia podemos encontrar ejemplos de la realización de censos, registros de producción, registros de transacciones comerciales, registro de impuestos, registros de bienes y propiedades, catastros, entre otros eventos que demuestran que la necesidad de registrar, organizar y presentar los datos para realizar toma de decisiones ha ido evolucionando, haciéndose cada vez más complejo a medida que transcurre el tiempo.

No fue sino hasta el siglo XIX cuando, gracias a los trabajos de Galton y Pearson, la estadística entra en escena como disciplina al diferenciarse la estadística deductiva de la inductiva y al generalizarse un método para estudiar fenómenos tanto de las ciencias naturales como de las sociales.

Una etapa muy importante se debe a R.A. Fisher, quien en el año de 1925 publica *Statistical Methods for Research Workers* y en el año 1935 *The Design of Experiments*, en estas obras presentó sus trabajos acerca de las aplicaciones del análisis de varianza y propone por primera vez el método para el procesamiento de datos de datos estadísticos que sigue en uso actualmente.

Con la aparición de las computadoras a mitad del siglo XX, surgen los programas de análisis estadístico, con los cuales la metodología estadística hace uso del poder de procesamiento de las computadoras para aplicar técnicas de procesamiento automatizado a grandes volúmenes de datos, con lo que el método

estadístico empieza a considerarse un proceso iterativo, a través del cual es posible determinar el modelo de análisis ideal para cada conjunto de datos. Al facilitarse el acceso a las computadoras personales y a los programas especializados en análisis estadístico (como SPSS, SAS, STATA, etc.), el uso de herramientas estadísticas se generaliza, los tiempos de análisis se reducen, aparecen continuamente herramientas accesibles tanto a profesionales del ramo como a público interesado en las diversas aplicaciones de la estadística.

Definiendo muy grosso modo la estadística, podemos afirmar que es la ciencia que se encarga de la colecta, acopio, organización, procesamiento, presentación e interpretación de datos obtenidos de una población o muestra con el objeto de describir su composición, entender su estructura y emitir predicciones considerando su comportamiento previo con cierto grado de certeza.

De la definición anterior se derivan una serie de conceptos importantes que es necesario abordar. Una POBLACIÓN es la totalidad de individuos u objetos que poseen una característica o participan en un fenómeno que nos interesa describir, analizar, entender o predecir. En múltiples ocasiones, también se le conoce como UNIVERSO de estudio.

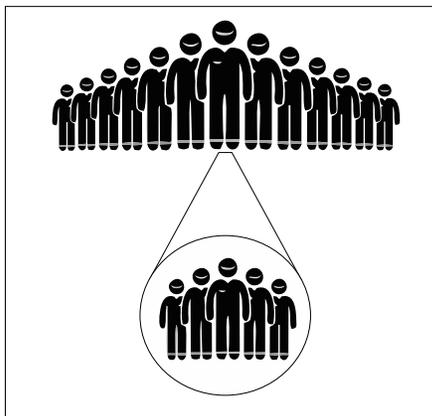
Así pues, si se desea determinar la estructura de edades de los habitantes de la Ciudad de México, la población estaría constituida por la totalidad de personas vivas que tienen su residencia en la Ciudad de México, es decir, todos los seres humanos que presentan los dos requisitos de interés: un tiempo de vida transcurrido desde su nacimiento y su lugar de residencia en la Ciudad de México.

Realizar un estudio en la totalidad de individuos de una población puede ser un proceso muy largo y complicado, dado el elevado número de elementos que las constituyen, el área geográfica en la que se

Documento educativo

distribuyen y el proceso que debe realizarse para obtener la información. Para evitar que los resultados sean poco pertinentes dado que pueden llevar mucho tiempo y recursos, muchos estudios se realizan en un subconjunto de la población que está conformado por individuos en los que las características de estudio presentan, en un escenario ideal, la misma distribución que en la población. Este subconjunto se conoce como MUESTRA.

Figura 1 Población y muestra



A fin de que la muestra presente, lo más apegado posible, las características de la población se han desarrollado técnicas y estrategias de selección de los individuos y del número a considerar de los mismos dentro del estudio, que en conjunto conforman la Teoría de Muestreo; entre ellas, la selección aleatoria (al azar) de los individuos, el cálculo del tamaño de muestra, entre otros.

Un resumen de las características de la población y de la muestra se presentan en el siguiente cuadro:

Tabla 1 Características de las poblaciones y las muestras

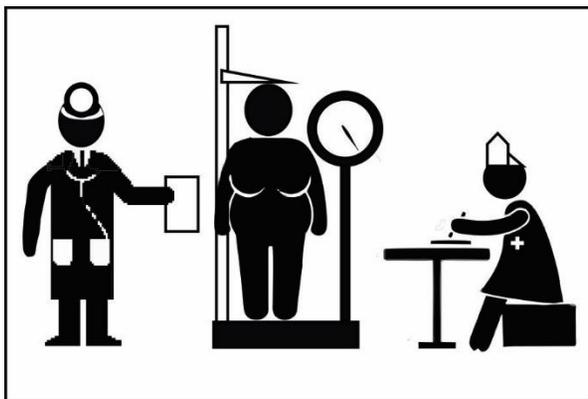
	Población	Muestra
Definición	Totalidad de elementos/individuos que poseen la(s) característica(s) motivo del estudio.	Selección de un subconjunto de la población que presenta la(s) característica(s) sujeto de estudio.
Características	Pueden clasificarse de acuerdo a la cantidad de individuos que la conforman (poblaciones finitas o infinitas). Posee características que pueden ser medidas, estimadas o valoradas de forma cualitativa. Siempre se encuentra ubicada en un espacio y tiempo determinados por el investigador en función de los objetivos del estudio que se realiza.	Forma parte de la población. Refleja las características de la población global. Los elementos que la conforman se eligen de forma aleatoria del total de individuos de la población. Debe ser representativa. Está asociada a un grado de confianza y un porcentaje de error dado que siempre es posible dejar fuera algún sector de la población. Está ubicada en el mismo espacio y tiempo que la población de la que se deriva.
Objetivos	Analizar los datos recabados referentes a las características comunes que comparten los elementos con diversos propósitos.	Estudiar el comportamiento, características, gustos o propiedades de una parte representativa de la población.

Documento educativo

	Población	Muestra
Ejemplos	Las personas que habitan una ciudad. La cantidad de vehículos en una ciudad. Los árboles que crecen en una región definida.	Para el estudio del desempeño de los estudiantes de cinco universidades de una ciudad en una materia específica, se toma como muestra a 500 estudiantes aleatoriamente (100 de cada institución) que estén cursando el mismo nivel para que la muestra sea representativa.

Las características que se estudian dentro de las poblaciones y muestras, al tratarse de propiedades, características o procesos que pueden adquirir diferentes valores en los elementos que conforman el universo de estudio, reciben el nombre genérico de **VARIABLES**.

Figura 2 Variables biométricas



Una variable, de acuerdo al tipo de valor que puede adquirir, se clasifican en cuantitativas y cualitativas. Las **VARIABLES CUANTITATIVAS** están conformadas por aquellas características que pueden ser medidas con una escala universal, cuyos valores son equidistantes uno de otro (por ejemplo,

centímetros, grados, milímetros de mercurio, mililitros) o contadas utilizando números naturales (valores enteros), por ejemplo: número de hijos, años cumplidos, número de cigarrillos que fuma por unidad de tiempo, entre otros.

A su vez, las variables cuantitativas pueden ser **CONTINUAS**, si adquieren un valor **DENTRO DE UN INTERVALO** que puede ser **EXPRESADO** con una escala que presenta una infinidad de valores intermedios entre dos puntos de medición; este valor que pueden registrar un mayor número de subunidades al incrementar la precisión de la medición de la variable; por ejemplo, estatura de un metro setenta y dos centímetros, dos milímetros, tres micrómetros. Por otra parte, pueden ser **DISCRETAS** en caso de que adquieran un valor que pueda contarse o expresarse a través de un número entero, por ejemplo, número de accidentes, cantidad de sillas en un aula, número de alumnos aprobados, profesores con aula virtual, etc.

Las **VARIABLES CUALITATIVAS** constituyen características que no pueden ser medidas utilizando una escala universal con valores equidistantes y que solo pueden ser percibidas o apreciadas por el investigador; por ejemplo, el color de ojos, el estado civil, la religión que profesan los participantes, el grado de mejoría ante un tratamiento, el nivel de dolor que se percibe, entre otros.

En el caso de las variables **CUALITATIVAS**, podemos tener variables cuyo valor constituya un valor que corresponda a una característica percibida o categoría sin una componente de orden, en cuyo caso tendríamos una **VARIABLE NOMINAL** (Proveniente del latín *nominālis*, relativo al nombre), como por ejemplo el color, el departamento al que se encuentra adscrito un trabajador, la escuela de procedencia, la licenciatura que estudia un alumno, entre otras.

Documento educativo

En caso de tener una característica que pueda ser expresada utilizando una escala, o que pueda adquirir un valor cuyos elementos tienen una componente de orden, esa característica constituirá una VARIABLE ORDINAL (Proveniente del latín *ordinālis*, relativo al orden), por ejemplo, lugar obtenido en una competencia (primero, segundo, tercero), calificación obtenida en un examen (No acreditado, suficiente, Bueno, Muy bueno, Excelente), grado de mejoría ante un tratamiento (ninguno, bajo, regular, alto muy alto, entre otros).

En la actualidad, la estadística se divide en tres ramas principales, considerando su campo de acción y objetivos, los cuales son:

I.- Estadística Descriptiva

La estadística descriptiva es la parte de la estadística que se dedica a describir el comportamiento de las variables dentro de una muestra o población y a presentar de manera resumida (en tablas o gráficas) de forma cuantitativa (medible) las características estudiadas, con base en el acopio de datos obtenidos de un proceso de colecta de información.

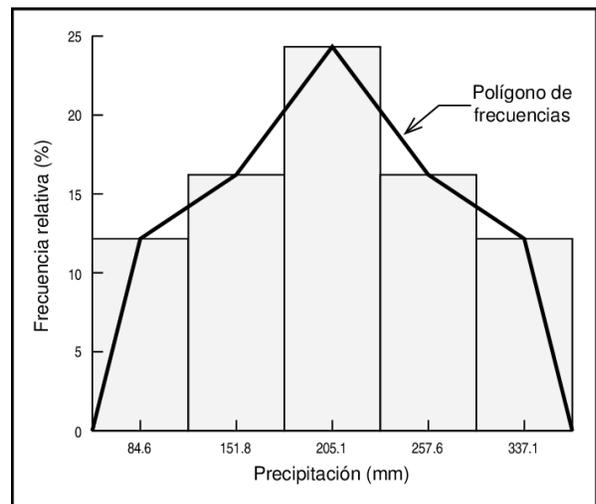
La estadística descriptiva se encarga de coleccionar, procesar, resumir y presentar datos provenientes de una población o una muestra estadística (subconjunto REPRESENTATIVO de elementos extraídos de la población) en función de parámetros O ESTADÍSTICOS obtenidos de su análisis; es importante señalar que esta parte de la estadística no busca relaciones entre variables o explicaciones acerca de su comportamiento, simplemente describe la forma en que los datos se presentan en los elementos o individuos que conforman la muestra o población.

Para que una muestra sea REPRESENTATIVA de una población, es decir, que los resultados del estudio puedan ser EXTRAPOLABLES a la totalidad de la

POBLACIÓN, debe cumplir con una serie de criterios:

- a) Los elementos/individuos deben ser electos preferentemente de forma aleatoria.
- b) Los elementos/individuos deben representar las características de estudio dentro de la población
- c) Debe considerar todos los diferentes perfiles de elementos/individuos de la población
- d) Debe tener un número de elementos/individuos, acorde al tamaño de la población

Figura 3 Representación de frecuencias en estadística descriptiva



Para realizar una descripción adecuada, la estadística descriptiva utiliza lo que conocemos como PARÁMETROS y ESTADÍSTICOS; es decir, un conjunto de medidas que sirven para describir un conjunto de datos, entre los más importantes están las medidas de tendencia central, las medidas de variabilidad o dispersión, las medidas de posición y las medidas de forma.

Un PARÁMETRO es un valor numérico calculado a partir de datos obtenidos DIRECTAMENTE de una POBLACIÓN, mientras que un ESTADÍSTICO es

Documento educativo

un valor numérico calculado a partir de datos obtenidos de una MUESTRA.

Dentro de las medidas de tendencia central, que son valores que estiman la forma en que los datos se concentran en torno de un valor de referencia (usualmente al CENTRO del conjunto de datos), se pueden destacar la media, la mediana y la moda; en el caso de las medidas de dispersión, que son valores que estiman la dispersión existente dentro de un conjunto de datos, los más usuales son el rango, el coeficiente de variación, la varianza y la desviación estándar; de la misma manera, dentro de las medidas de posición, que estiman la ubicación de un valor dentro de la distribución de datos, se encuentran los percentiles, y, finalmente, existen las medidas de forma, que indican la apariencia de la distribución de los datos, entre las que podemos destacar la asimetría y la curtosis.

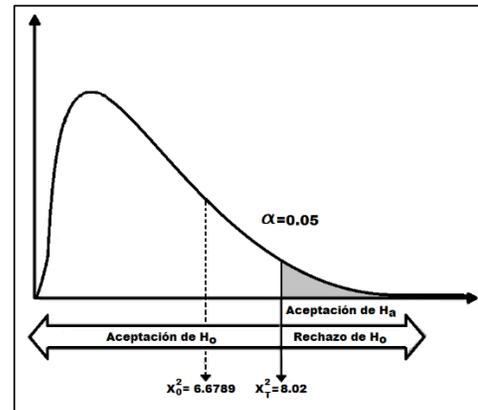
Cuando se realiza un estudio de tipo estadístico, la estadística descriptiva suele ser el primer paso del análisis, ya que proporciona una serie de datos que permiten al investigador entender las características de la población o muestra en estudio; los cuales se presentan usualmente a través de tablas y gráficos. La estadística Descriptiva constituye la base de casi cualquier análisis cuantitativo (medible) de datos.

II.- Estadística Inferencial o inferencia estadística

Mientras la estadística descriptiva busca representar la forma en la que las variables se presentan dentro de una población o muestra, la inferencia estadística o estadística inferencial busca identificar asociaciones causales o relaciones entre diferentes variables de estudio, auxiliándose del uso de herramientas de inferencia e inducción. Esta rama de la estadística busca, a partir de una serie de parámetros, deducir el comportamiento de una población estudiada a partir de datos obtenidos a partir de la misma de forma directa o mediante una o

varias muestras; esto quiere decir que no solo recolecta y resume los datos, sino que busca explicar ciertas dinámicas a través de la interacción de las variables estudiadas con base en los datos obtenidos.

Figura 4 Prueba de contraste en inferencia estadística



Es muy importante señalar que, para que la estadística inferencial permita obtener conclusiones correctas de un análisis estadístico, es necesario que el investigador previamente haya realizado un buen análisis y resumen de los datos a través de estadística descriptiva. De aquí la importancia de que se realicen ambos estudios de manera adecuada, ya que la estadística descriptiva provee el insumo necesario para que la estadística inferencial cumpla su cometido.

En el terreno de la inferencia estadística, las conclusiones obtenidas tienen siempre un componente de aleatoriedad, en los que la variabilidad y el azar juegan un papel fundamental; la presencia de variabilidad implica que cada una de las pruebas realizadas esté asociada a un índice de confianza, de significancia y de error, los cuales deben reportarse junto con los resultados del estudio, conceptos que se retomaran y abordarán adecuadamente más adelante en esta serie.

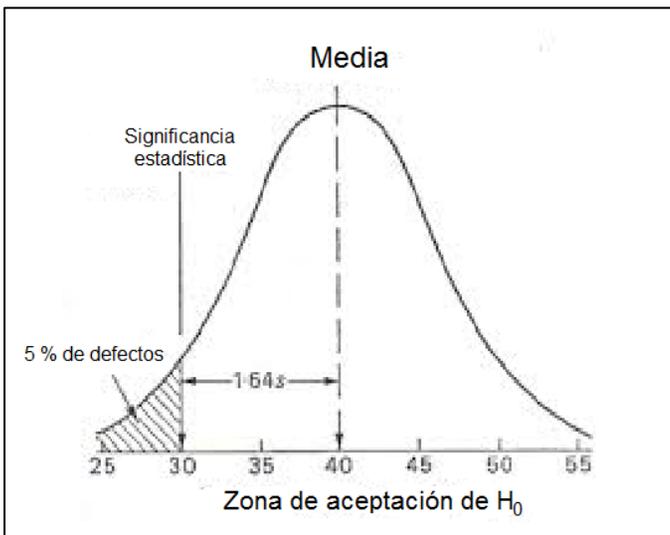
Documento educativo

Dependiendo de las características propias de la población bajo estudio, la estadística inferencial se divide en:

a) Estadística paramétrica

Este tipo de estadística comprende procedimientos estadísticos basados en la distribución CONOCIDA de los datos reales, los cuales se determinan mediante un número finito de parámetros estadísticos (número que resume la cantidad de datos derivados de una variable estadística), estos parámetros son el insumo necesario para que se determine la asociación, relación o independencia de las variables que se consideran en el estudio.

Figura 5 Ejemplo de prueba paramétrica

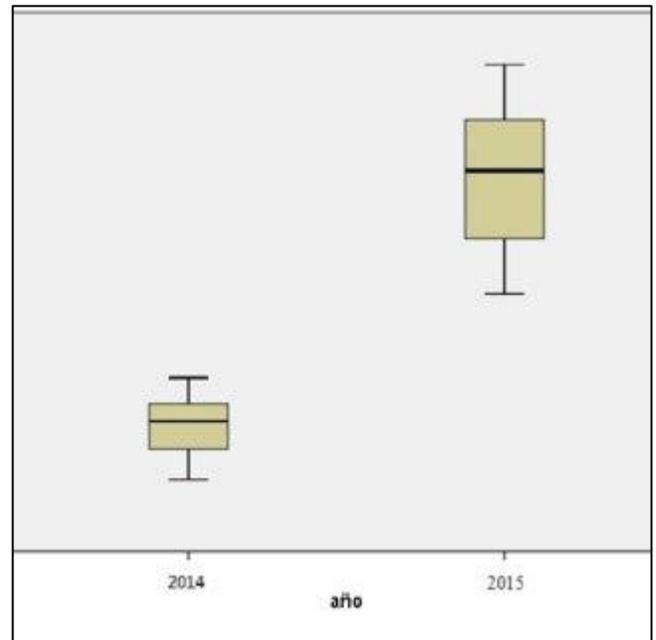


Para que dentro de un estudio se apliquen procedimientos de tipo PARAMÉTRICO, se requiere conocer previamente la forma en que los datos se distribuyen dentro de la población, así como la forma resultante de la gráfica de distribución de la población estudiada. Si se desconoce en su totalidad la distribución que siguen los datos obtenidos, se debe utilizar un procedimiento no paramétrico.

b) Estadística no paramétrica

En esta rama de la estadística inferencial, se incluyen los procedimientos aplicados en pruebas y modelos estadísticos en los que se desconoce la forma en que se distribuyen los datos al interior de la población y, por lo tanto, la distribución de datos no se ajusta a los llamados criterios paramétricos. Al ser los datos estudiados independientes a una distribución teórica, sus técnicas pueden ser aplicadas efectivamente a un gran conjunto de procesos, de los que se carece de información previa o que su población es tan increíblemente grande y compleja, que no puede ser estudiada directamente por su amplia distribución geográfica o su enorme número.

Figura 6 Resultados de un análisis no paramétrico



Cuando en un estudio se desconoce si los datos existentes se ajustan a una distribución teórica conocida, el investigador debe elegir, sin duda, técnicas pertenecientes a la estadística no paramétrica, de manera que puede ser un paso previo al procedimiento paramétrico.

Documento educativo

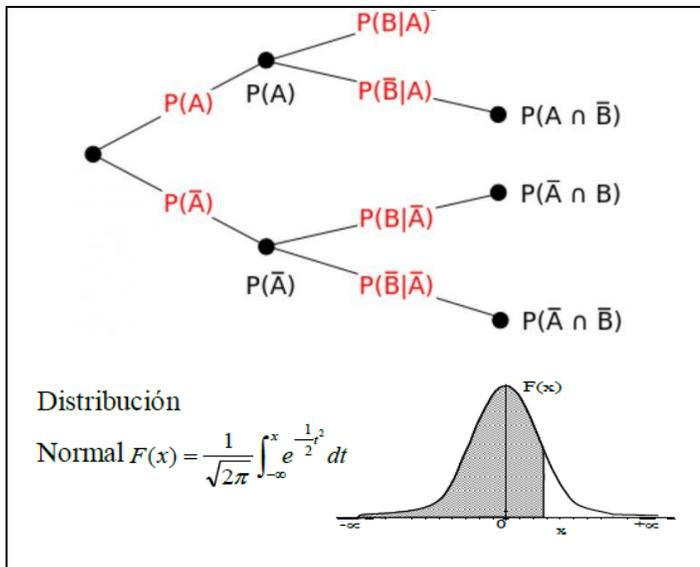
De la misma manera que en el caso de la estadística paramétrica, las posibilidades de error se disminuyen mediante el uso de tamaños muestrales adecuados a las características y tamaño de la muestra/población bajo estudio.

3- Estadística Matemática

Algunos autores diferencian la estadística matemática como una disciplina diferente a las otras dos ramas de la estadística.

Esta disciplina consiste en un abordaje teórico del estudio de los datos, en el cual se utilizan la teoría de la probabilidad (rama de las matemáticas que estudia los fenómenos aleatorios), el álgebra, la geometría analítica y otras ramas de las matemáticas.

Figura 7 Estadística matemática, fundamentos teóricos de la estadística aplicada



La estadística matemática consiste en la obtención de información a partir de los datos y utiliza técnicas matemáticas tales como: análisis matemático, álgebra lineal, análisis estocástico, ecuaciones diferenciales, geometría analítica, etc. Proveyendo a los investigadores de campo de herramientas y técnicas de análisis que les permitan obtener información confiable y de modelos de predicción

acordes con las características de las poblaciones, cuyo índice de confianza, significancia y error puedan ser determinados mediante la aplicación de procedimientos confiables. Así, la estadística matemática influye en la estadística aplicada y provee a las otras ramas de herramientas, técnicas y procedimientos adecuados para la obtención de conclusiones basadas en datos reales y un alto grado de confianza.

Próximo documento educativo:

Estadística descriptiva (I): variables y frecuencias.

Referencias

Anderson, D. R., Sweeney, D. y Williams, T. A. (1999). *Estadística para la Administración y Economía*. México DF, México: International Thompson Editores.

Departamento de Didáctica de la Matemática. (2011). *Estadística con proyectos*. (C. Batanero, & C. Díaz, Eds.) Granada, España: Facultad de Ciencias de la Educación, Universidad de Granada.

García Pérez, A. (2008). *Estadística aplicada: conceptos básicos (2a edición ed.)*. Madrid, España: Educación permanente / Universidad Nacional de Educación a Distancia.

Hanlle, V. (2011). *Proyecto de estudio de las pausas activas en el Clima Laboral y su influencia e impacto para la motivación y satisfacción física de los empleados de Premex Ecuador en la Ciudad de Quito*. Tesis. Universidad de las Américas, Quito.

Jaraiseh, N. (2015). *Estrés Laboral y Síndrome de Burnout: Pausas activas como método de afrontamiento*. Tesis. Universidad Internacioanl SEK, Quito.

Kazmier, L. J., Díaz Mata, A., y Eslava Gómez, G. (1991). *Estadística Aplicada a Administración y Economía*. Naucálpan, Estado de méxicoméxico, Atlacomulco, México: mcgraw Hill.

Documento educativo

*Pérez López, C. (1999). Control estadístico de la calidad.
Madrid, España: Alfa Omega.*

*Wackerly, D. D., Mendenhall III, W., y Scheaffer, R. (2010).
Estadística Matemática con aplicaciones. México,
D.F., México: Cengage Learning Editores, S.A.*

Declaración de conflicto de intereses

Los autores declaran no tener ningún interés comercial o asociativo que represente un conflicto de intereses en relación con el trabajo presentado.

Obra protegida con una licencia Creative Commons

