

La estadística. Una herramienta en la investigación científica aplicada a la educación

The Statistics. A toolin scientific research applied to education

Gerardo R. Lara Morel

laragerardo@gmail.com

Universidad Católica Andrés Bello. Caracas. Venezuela

RESUMEN

La preocupación de los investigadores en presentar sus resultados de forma clara a los lectores y público en general los ha llevado a utilizar recursos de otras ciencias. Este es el caso, la Estadística como medio para abordar la interpretación y el análisis de los datos. Este artículo no recorrerá la estadística en todos sus contenidos sino como a través de sus definiciones se logra presentar la información recabada de manera sistemática.

Palabras clave: estadística, variables, media, moda, mediana, variabilidad, rango, diagrama de caja, tabla e histograma.

ABSTRACT

Researchers' concern in presenting their results clearly to readers and the general public has led them to use resources from other sciences. This is the case, Statistics as a means to approach the interpretation and analysis of data. This article will not go through the statistics in all its contents but as through its definitions it is possible to present the information gathered in a systematic way.

Key words: statistics, variables, mean, mode, median, variability, range, box diagram, table and histogram.

INTRODUCCIÓN

La rapidez con la cual se desarrolla la ciencia invita a los investigadores a una constante actualización o adaptación a la manera de abordar las distintas problemáticas y así facilitar el manejo de la información y la divulgación de las mismas. Es por eso que de una manera amena se presentan los conceptos de la

Estadística Descriptiva y como presentarlo a partir de las variables en cuadros y gráficos, sin perder de vista el rigor científico de la misma.

La estadística y el resto de las ciencias.

La estadística nace en el siglo II antes de Cristo con los Censos Chinos y/o con el control de siembra del Antiguo Imperio Egipcio. O más reciente cuando Achenwell, Gottfried en 1749 acuña la definición de estadística, como lo expresa la Buchanax (s/f), dando muestra de su aplicabilidad en casi todas las ciencias existentes actualmente.

Hoy en día la estadística es usada en forma directa o indirecta por todas las ciencias y más aún ha desarrollado métodos propios para cada una de ellas como un recurso inseparable para representar las características de un determinado proceso o experimentos, así como la búsqueda de respuesta a nuevas preguntas o incertidumbres encontrada en un determinado hecho o experimento.

La estadística, también, pasa a ser una ciencia básica cuyo objetivo principal es el procesamiento y análisis de grandes volúmenes de datos, resumiéndolos en tablas, gráficos e indicadores (estadísticos), que permiten la fácil comprensión de las características concernientes al fenómeno estudiado.

Ramas de la estadística

Pestaña de Martínez (2002) expone que la estadística está clasificada en dos grandes ramas como son: 1.- *Estadística descriptiva o deductiva*, que se dedica a los métodos de recolección, descripción, visualización y resumen de datos originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numérica o gráficamente. Ejemplos básicos de descriptores numéricos son: la media y la desviación estándar. Algunos ejemplos gráficos son: histograma, pirámide poblacional, clúster, entre otros y, 2.- *Estadística inferencial o inductiva*, que se dedica a la generación de los modelos, inferencias y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las

observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población bajo estudio.

Estas inferencias pueden tomar la forma de respuestas a preguntas si/no (prueba de hipótesis), estimaciones de características numéricas (estimación), pronósticos de futuras observaciones, descripciones de asociación (correlación) o modelamiento de relaciones entre variables (análisis de regresión). Otras técnicas de modelamiento incluyen ANOVA, series de tiempo y minería de datos.

Ambas ramas (descriptiva e inferencial) comprenden la estadística aplicada. Hay también una disciplina llamada estadística matemática, la cual se refiere a las bases teóricas de la materia. La palabra estadísticas también se refiere al resultado de aplicar un algoritmo estadístico a un conjunto de datos, como en estadísticas económicas. Esta trabaja sobre datos y dependiendo el tipo de dato ella presenta sus opciones para facilitar la información a dar a conocer a un público mayor y que pueda ser replicado o verificado en condiciones similares. Es por ello que debemos tener claros los diferentes tipos datos o variables estadística observables y sus posibles usos para presentar los mismos.

De las Variables a los Datos

Los investigadores llaman variables a los constructos utilizado en sus investigaciones y las mismas definen las propiedades que estudian. Algunas de éstas: edad, grado de instrucción, género, estatura, presión arterial, estrés y aprovechamiento, entre muchas otras. A partir de aquí se puede decir que variable es una propiedad del objeto o sujeto, que adquiere valores distintos.

Tipos o Clases de Variables

Variables Cualitativas. Son aquellas que describen los atributos del sujeto u objeto de estudio como lo son color, textura, genero, entre otras. Se ha de observar que entre sus categoría no existe relación numérica alguna.

Variables Cuantitativas. Admiten una escala numérica de medición y son susceptibles de cuantificación como lo son la edad, número de hijos, nivel de estrés, el aprendizaje, etc.

Las variables cuantitativas pueden ser:

Discretas, que no pueden adoptar valores intermedios entre un valor y otro.

Continua, cuando puede tomar cualquier valor entre dos valores dados.

Basándose en la tabla presentada por Triola (2004) y adaptando a través del siguiente cuadro explica cómo interpretar las variables y su nivel de medición.

Cuadro 1
Las variables y su nivel de medición

Nivel de Medición	Comparación de valores	Relación	Ejemplos
Nominal	Valores iguales o diferentes	$a=b$. $a \neq b$	Genero, color del cabello, religión
Ordinal	Un valor es mayor o menor que otro.	$a > b$ $a < b$	Grado de instrucción, estado civil Posiciones en una jerarquía.
Intervalo	Diferencia de valores pueden compararse.	$(a-b) < (c-d)$	Coficiente intelectual
Razón	Relación de valores puede compararse.	$a/b < c/d$	Masa corporal, peso de un cuerpo, Ingreso, duración de un matrimonio

Adaptada de Triola (2004)

Nivel de medición o escala de la variable

El investigador debería anticiparse y considerar en el proyecto de investigación, en particular en el tipo de instrumento a utilizar para la recolección de los datos (por ejemplo una escala, una ficha de ingreso, un cuestionario, etc.), esto porque la estadística requiere variables que reúnan ciertos niveles de medición. Mientras mayor sea el nivel de medición más herramientas estadísticas tendrá el investigador para abordar la descripción y el análisis estadísticos de la variable.

Las variables Nominales son las primeras en la escala seguidas de las Ordinales, donde se puede ordenar pero no existe medida justa entre uno rango y el siguiente. Las de Intervalos o de Razón son aquellas que pueden relacionarse en forma unívoca con algún subconjunto de los números reales (R) y por lo tanto son descritas de una forma rica, porque lo que se verifique en R vale en dichas variables y son propensas a verificación.

Cuadro 2
Nivel de medición o escala de la variable

Nivel de Medición	Descripción y presentación	Prueba de hipótesis	Medida de asociación o correlación*
Nominal	<ul style="list-style-type: none"> • Tabla de Frecuencia • Gráfica de Sectores • Moda • Coeficiente cualitativo de variabilidad 	<ul style="list-style-type: none"> • Chi cuadrado • Fisher¹ • Mc Nemar¹ 	<ul style="list-style-type: none"> • V de Cramer • Coeficiente de contingencia (C) • Phi (ϕ) • Lambda (λ) • Tau(T)
Ordinal	<ul style="list-style-type: none"> • Tabla de Frecuencia • Histograma • Moda • Mediana • Rango 	Grupos Independiente: <ul style="list-style-type: none"> • U de Mann-Whitney • Kruskal- Wallis Grupos relacionados: <ul style="list-style-type: none"> • Wilcoxon t-test • Fredman Anova 	<ul style="list-style-type: none"> • Rho de Spearman (r_s) • d de Sommers (d) • Gamma (γ)
Intervalo o Razón	<ul style="list-style-type: none"> • Tabla de frecuencia • Polígono de frecuencia • Histograma • Moda • Mediana • Media aritmética • Rango • Rango intercuartil • Varianza • Desviación Típica • Coeficiente de variabilidad 	Grupos Independiente: <ul style="list-style-type: none"> • T de Student • Anova factorial Grupos correlacionales: <ul style="list-style-type: none"> • T de Student • Anova factorial 	2 variables: <ul style="list-style-type: none"> • r Pearson Más de 2 variables: <ul style="list-style-type: none"> • R Múltiple (R^2) Otros: <ul style="list-style-type: none"> • Eta y Eta² (η) • Punto biseral (Y_{pb})

Fuente: Landero y González (2006)

¹ Requiere variables dicotómicas.

*Esta depende del nivel de medida de las variables (X, Y) y, en su caso del número de filas y columnas de la tabla.

A partir de la segunda columna del cuadro anterior se describirá en forma ordenada las definiciones estadísticas de cada término y sus propiedades sin el uso de fórmulas engorrosas, siempre que se necesiten se usará el programa Statistical Package for the Social Sciences, y se denotará por sus siglas SPSS, versión 19 o superior.

1.- **La tabla de frecuencias** (o distribución de frecuencias) es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales.

Ejemplo:

Cuadro 3
Publicaciones de Tesis por año

Fecha	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
1994	1	1,4	1,4	1,4
1996	1	1,4	1,4	2,8
1997	1	1,4	1,4	4,2
2000	2	2,8	2,8	7,0
2001	1	1,4	1,4	8,5
2003	1	1,4	1,4	9,9
2004	7	9,9	9,9	19,7
2005	12	16,9	16,9	36,6
2006	8	11,3	11,3	47,9
2007	7	9,9	9,9	57,7
2008	9	12,7	12,7	70,4
2009	8	11,3	11,3	81,7
2010	5	7,0	7,0	88,7
2011	5	7,0	7,0	95,8
2012	1	1,4	1,4	97,2
2013	1	1,4	1,4	98,6
2014	1	1,4	1,4	100,0
	71	100,0	100,0	

Cuando los datos son abundantes suelen agruparse para facilitar la comprensión de la tabla de frecuencia y se denomina Tabla de Frecuencia de datos

agrupados, aun cuando para el SPSS no hay dificultad de tratarlo como tabla simple (sin agrupar).

2.- Gráfica de Sectores. Un diagrama de sectores es un gráfico que consiste en un círculo dividido en sectores de amplitud proporcional a la frecuencia de cada valor. Se utiliza con datos cualitativos y cuantitativos.

Ejemplo:

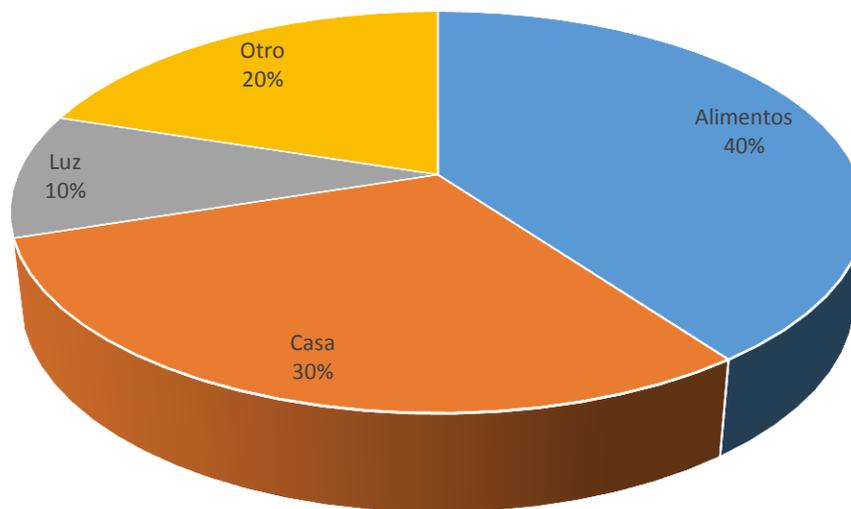


Gráfico 1
Gastos del ingreso familiar

3.- **Moda o Modo**, es el valor con mayor frecuencia en una distribución de datos.

Ejemplo:

Si observa el cuadro 3 la moda es el año 2005 con una frecuencia de 12, de igual forma en el cuadro 4 el género masculino con una frecuencia de 111.

4.- **El índice de variación cualitativa (IVC)**, El IVC expresa el grado en que los casos están dispersos en las diferentes categorías de la variable,

alcanzando su máximo ($IVC = 1$) en el caso en que las frecuencias relativas sean iguales para todas las categorías de la variable (caso que se corresponde al de una variable con una distribución uniforme). El IVC sería igual a 0 cuando la frecuencia relativa de una categoría de la variable fuese igual a 1, esto es, el caso en que todos los casos tuviesen el mismo valor observado en la variable (dispersión nula). El índice de variación cualitativa (IVC)

- Se obtiene a través de la siguiente fórmula, donde k es el número de categorías de la variable y p_i la frecuencia relativa asociada a cada una de ellas:

$$IVC = \frac{1 - \sum_{i=1}^K p_i^2}{\left(\frac{k-1}{k}\right)}$$

Ejemplo: Sea la variable “Religión que se profesa” [Codificación: 0: Católica; 1: Protestante; 2: Otra; 3: Ninguna], de la que se han obtenido datos para una muestra de 50 personas, cuya distribución de frecuencias se muestra a continuación

Cuadro 4
Religión que se profesa

X_i	Frecuencia absoluta	Frecuencia Relativa
0	12	0,24
1	10	0,2
2	10	0,2
3	18	0,36
	50	1,00

Al sustituir en la expresión anterior, las frecuencias relativas, se obtiene.

$$IVC = \frac{1 - (0,24^2 + 0,2^2 + 0,2^2 + 0,36^2)}{\left(\frac{4 - 1}{4}\right)} = 0,98$$

5.- **Histograma**, es una representación gráfica de una variable en forma de barras. Se utilizan para variables continuas o para variables discretas, con un gran número de datos, y que se han agrupado en clases. En el eje abscisas se construyen unos rectángulos que tienen por base la amplitud del intervalo, y por altura, la frecuencia absoluta de cada intervalo. La superficie de cada barra es proporcional a la frecuencia de los valores representados.

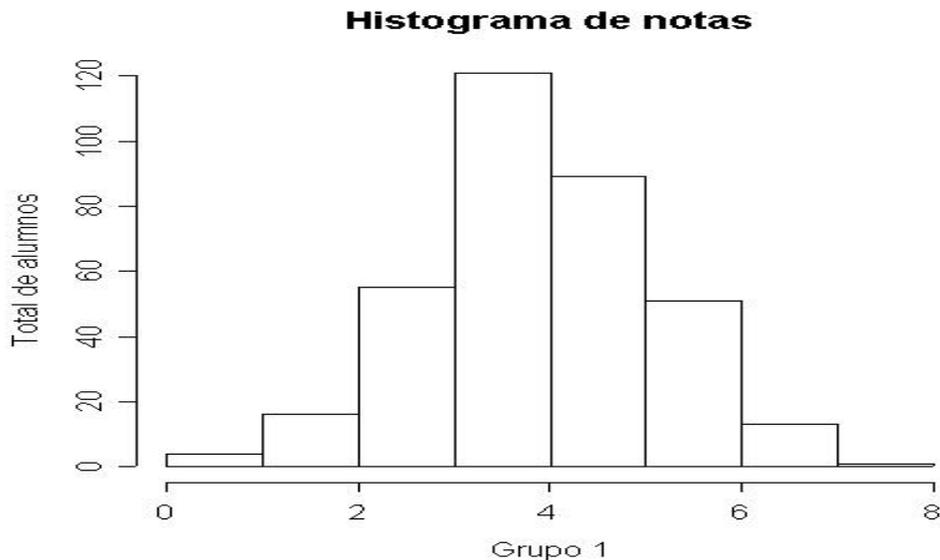


Gráfico 2
Promedios de notas de estudiantes universitarios

6.- **Mediana**¹, es el número central de un grupo de números ordenados por tamaño. Si la cantidad de términos es par, la mediana es el promedio de los dos números centrales.

Ejemplo:

¹ Al final se presentará un ejemplo donde se obtienen todos los resultados, con el uso del SPSS

Si observa el cuadro 3 la mediana es el dato que ocupa la posición 36, porque 35 de los datos están por debajo de dicha posición y 35 por encima, después de ordenarlos.

7.- **Rango**, es el intervalo entre el valor máximo y el valor mínimo; por ello, comparte unidades con los datos. Permite obtener una idea de la dispersión de los datos, cuanto mayor es el rango, más dispersos están los datos de un conjunto.

Ejemplo:

Si observa el cuadro 3 el Rango es la diferencia entre 2014 y 1994, será $2014-1994=20$.

8.- **Media aritmética**, es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de datos.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Ejemplo:

Si tienes las siguientes edades de niños de un maternal; 4, 5, 4, 5, 5, 4, 6, 6, 5, 5, 4; su media aritmética será:

$$\bar{X} = \frac{4 + 4 + 4 + 4 + 5 + 5 + 5 + 5 + 5 + 6 + 6}{11} = 4,81$$

9.- **Rango intercuartil**, se le llama rango intercuartílico o rango intercuartil, a la diferencia entre el tercer y el primer cuartil de una distribución. Es una medida de la dispersión estadística.

Ejemplo:

Si tienes las siguientes edades de niños de un maternal; 4, 5, 4, 5, 5, 4, 6, 6, 5, 5, 4; debes ordenar para calcular el primer y tercer cuartil: 4 4 4 4 5 5 5 5 5 6 6 los datos subrayados representan los cuartiles (Primer cuartil=4; Segundo cuartil 5; Tercer cuartil 5), para el cálculo necesitamos sólo el tercer y el primer cuartil, es decir, $R_{Ic} = 5-4 = 1$.

10.- **Varianza**, es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística. Se representa por σ^2 y su

fórmula es: $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n}$

Ejemplo:

Calcular la varianza de la distribución: 9, 3, 8, 8, 9, 8, 18.

Deben calcular la media aritmética.

$$\bar{X} = \frac{3 + 8 + 8 + 8 + 9 + 9 + 18}{8} = 9$$

Ahora la varianza

$$\sigma^2 = \frac{(3 - 9)^2 + (8 - 9)^2 + (8 - 9)^2 + (8 - 9)^2 + (9 - 9)^2 + (9 - 9)^2 + (9 - 9)^2 + (18 - 9)^2}{8} = 15$$

11.- **Desviación Típica**, La desviación típica o desviación estándar (denotada con el símbolo σ , es una medida de dispersión para variables de razón (variables cuantitativas o cantidades racionales) y de intervalo. Se define como la raíz cuadrada de la varianza de la variable

$$\sigma = \sqrt{\sigma^2}$$

Ejemplo:

Calcular la Desviación típica de la distribución: 9, 3, 8, 8, 9, 8, 18.

Tomando el resultado del ejemplo anterior, tendrás

$$\sigma = \sqrt{15} = 3,873$$

12.- **Coefficiente de variación** es la relación entre la desviación típica de una muestra y su media. El coeficiente de variación se suele expresar en porcentajes: El coeficiente de variación permite comparar las dispersiones de dos distribuciones distintas, siempre que sus medias sean positivas.

$$C.V = \frac{\sigma}{\bar{X}}$$

Ejemplo

De los resultados de los dos últimos ejemplos, calcular el coeficiente de variación.

$$C.V = \frac{3,873}{9} = 0,4303$$

Generalmente se expresa en porcentaje

$$C.V = \frac{\sigma}{\bar{X}} \cdot 100 = \frac{3,873}{9} \cdot 100 = 43,03\%$$

Se utiliza para comparar muestra en distintas unidades, por ejemplo puntajes de un test y notas obtenidas en una prueba, el valor menor será el que presenta menor variabilidad.

13.- Medidas de Posición. Las medidas de posición dividen un conjunto de datos en grupos con el mismo número de individuos.

Para calcular las medidas de posición es necesario que los datos estén ordenados de menor a mayor.

Las medidas de posición son:

Cuartiles. Los cuartiles son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales.

Q_1 , Q_2 y Q_3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos.

Q_2 coincide con la mediana.

Buscamos el lugar que ocupa cada **cuartil** mediante la expresión

$$\frac{k \cdot N}{4}, k = 1, 2, 3$$

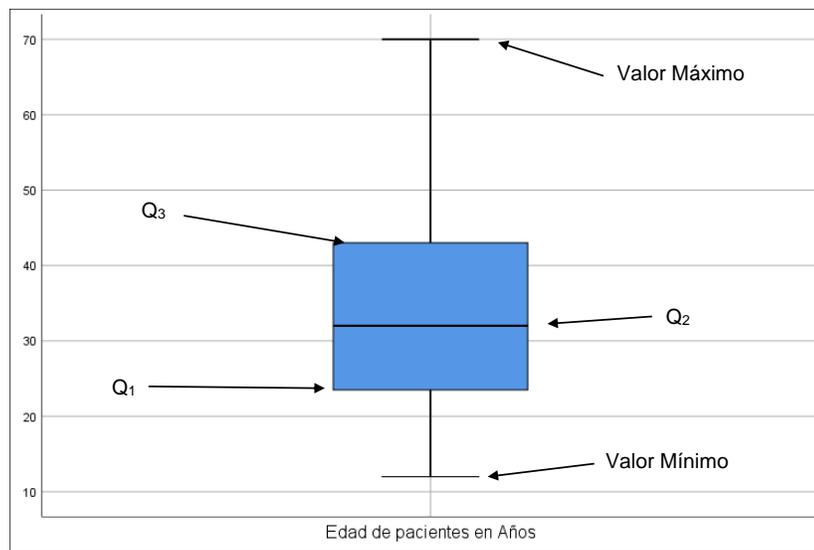


Gráfico 3
Diagrama de Caja

Es la representación de las medidas de posición de las variables tratada.

Deciles

Los deciles son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos.

D_5 coincide con la mediana.

Cálculo de los deciles

$$\frac{k \cdot N}{10}, \quad k = 1, 2, \dots, 9$$

Percentiles

Los percentiles son los 99 valores que dividen la serie de datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos.

P_{50} coincide con la mediana.

Cálculo de los percentiles

$$\frac{k \cdot N}{100}, \quad k = 1, 2, \dots, 99$$

Ejemplo

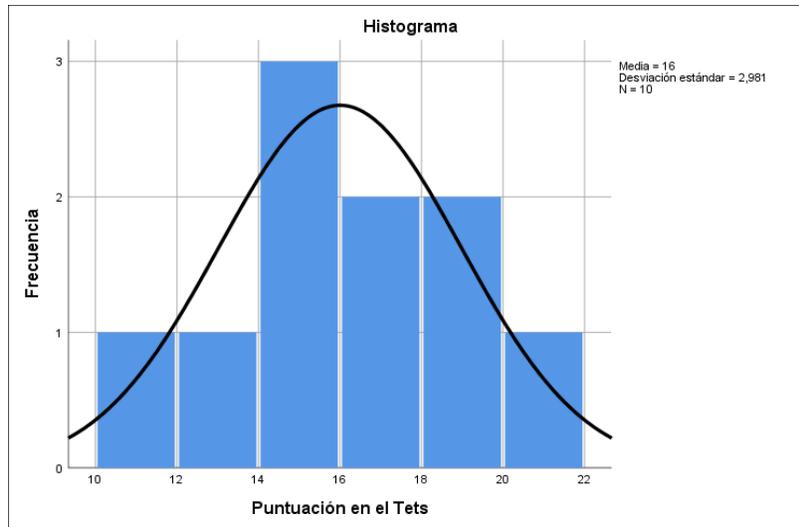


Gráfico 4 Histograma de la Temperatura en Grado Centrígrados.

Ejemplo.

Dado el siguiente conjunto de datos:

Alumno	Tets	Prueba
A	15	14
B	15	12
C	14	10
D	13	12
E	11	8
F	21	17
G	19	14
H	18	16
I	17	12
J	25	15

Cuadro 5

Estadísticos del Estudio.

		Estadísticos	
		Puntuación en el Tets	Calificación del curso
N	Válido	10	10
	Perdidos	0	0
Media		16,80	13,00
Error estándar de la media		1,306	,869
Mediana		16,00	13,00
Moda		15	12
Desv. Desviación		4,131	2,749
Varianza		17,067	7,556
Rango		14	9
Mínimo		11	8
Máximo		25	17
Percentiles	10	11,20	8,20
	20	13,20	10,40
	25	13,75	11,50
	40	15,00	12,00
	50	16,00	13,00
	75	19,50	15,25

Datos tomados de Hamdan (1994). Así lo presenta el programa SPSS después de cargar la información.

El Coeficiente de variación de debe calcular por medio de la fórmula.

$$CV_{Test} = \frac{4,131}{16,80} * 100 = 24,589\%$$

$$CV_{Calificación} = \frac{2,749}{13} * 100 = 21,146\%$$

Conclusiones

Se mostró a través de las definiciones de la columna 2 lo que los estadísticos denominan un análisis exploratorios de Datos Descriptivamente, el orden a seguir no es estricto, la experiencia y el uso de programas lo realizan de manera simultánea en un instante de tiempo breve.

El siguiente mapa muestra en su lado izquierdo lo desarrollado en el artículo.

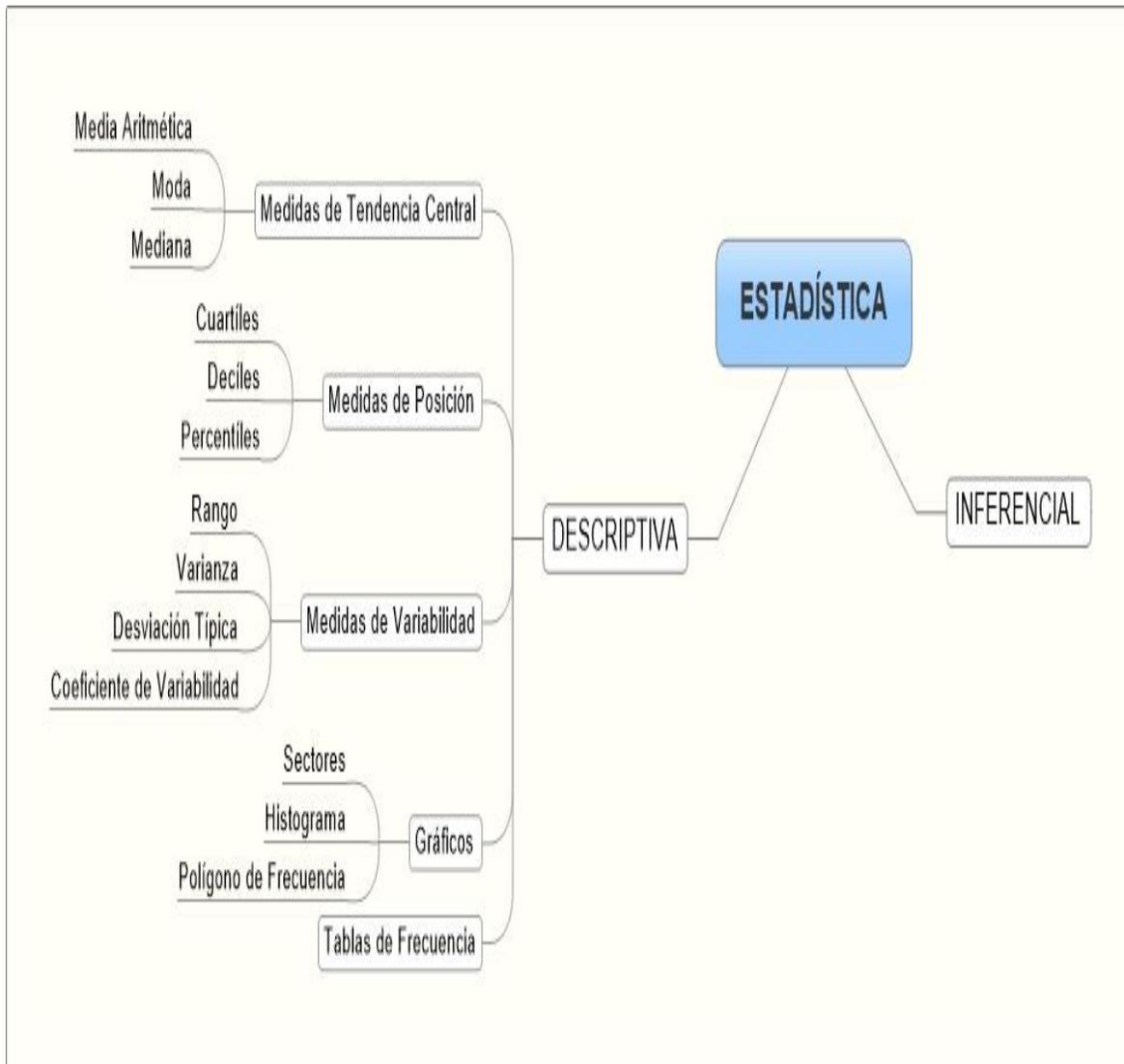


Gráfico 3
Conceptos desarrollados

REFERENCIAS

Buchanax, M. (s/f). Revista de España. Presented to The Library by professor Milton A. Buchanax of the department of Italiax and Spanish *Revista de España* Decimoquinto Año. Tomo LXXXVIII. Septiembre y Octubre. Madrid.

Landero, R. y González, M. (2006). *Estadística con SPSS y Metodología de la Investigación*. México: Trillas.

Pestaña de Martínez, P. (2002). *Estadística. Conceptos básicos, terminología y metodología de la estadística descriptiva*. Venezuela: Los Libros de El Nacional.

Triola, M. F (2004). *Estadística*. México: Pearson Addison Wesley.