



**Nombre del alumno:**

**Gomez Velasco Blanca Nabora**

**Nombre del profesor:**

**Ojeda Trujillo Juan**

**Nombre del trabajo:**

**Investigacion sobre estadistica descriptiva**

**Materia:**

**Estadística**

**Grado:**

**Primer cuatrimestre**

**Grupo:**

# INTRODUCCIÓN

La presente investigación se refiere al tema de la Estadística, que se pueden definir es la ciencia cuyo objeto es reunir una información para facilitar al hombre el estudio de datos masivos de individuos, grupos, series de hechos, etc. Y deducir de ello gracias al análisis de estos datos unos significados precisos.

La estadística descriptiva, recolectan y organizan datos que surgen como resultado, como también dice que la estadística descriptiva es una ciencia que analiza series de datos (por ejemplo, edad de una población, altura de los estudiantes de una escuela, temperatura en los meses de verano etc.) y trata de extraer conclusiones sobre el comportamiento de estas variables.

Las variables pueden ser dos tipos: como también veremos las aplicaciones descriptivas en las actividades de un administrador .como en donde encontramos la estadística con ejemplos para lograr una mayor comprensión sobre lo que es la estadística y la diferencia estadística descriptiva.

# Aplicación del proceso administrativo en los estudios estadísticos

**ESTUDIOS ESTADÍSTICO:** La estadística es la ciencia de los datos, la cual implica su recolección, clasificación, organización, análisis e interpretación, para la toma de decisiones frente a la incertidumbre.



El proceso administrativo es el flujo continuo e interrelacionado de las actividades de planeación, organización dirección y control, desarrolladas para lograr un objetivo común: aprovechar los recursos humanos, técnicos, materiales, y de cualquier otro tipo, con los que cuenta la organización para hacerla efectiva.

## *Las 4 etapas del proceso administrativo :*

**Planeacion** :Investigacion del entorno externo e interno,planteamiento de estrategias ,políticas y propósitos ,asi como de acciones a ejecutar en el corto ,medio y largo plazo.

### **EJEMPLO:**

Queremos abrir un empresa comenzamos con la planeacion y estructura de esta ...

Ayuda a llevar a cabo un estudio de mercado que permite tomar las mejores decisiones.

Nos sirve a buscar presupuestos y estrategias para lograr los objetivos deseados

**Organización:** La función de la organización es disponer y coordinar todos los recursos disponibles como son humanos, materiales y financieros.

**EJEMPLO:**

>Se organiza la información recabada para usos de la misma empresa.

>Se puede organizar información de diferentes maneras, mediante gráficos (tablas y tablas de frecuencia).

>Se puede organizar la información para mantener un control de la materia prima utilizada.

>La división del trabajo

>Recopilación de datos de la empresa

**Dirección:** Se encuentra la ejecución, los planes, la motivación, la comunicación y la supervisión para alcanzar las metas de la organización.

**EJEMPLO:**

>Nos ayuda a verificar los resultados obtenidos en las etapas anteriores.

>verificamos que la empresa va o no por buen camino.

>manejo de conflicto.

**CONTROL:** Es la función que se encarga de evaluar el desarrollo general de una empresa.

**EJEMPLO:**

- >Obtenemos los resultados finales.
- >Ganancias-perdidas gráficas y tabulaciones nos podremos dar cuenta de dichos resultados.
- >En caso de ganancia seguir con el mismo procedimiento y en caso de pérdidas cambios de estrategia y/o procedimientos.

## **Aplicación de la estadística descriptiva en las actividades del administrador**

Estadística descriptiva se ocupa de los métodos de recolección, descripción, visualización y resumen de datos originados a partir de los fenómenos en estudio. Los datos pueden ser resumidos numéricamente o gráficamente. Ejemplos básicos de descriptores numéricos son: la media y la desviación estándar. Resúmenes gráficos incluyen varios tipos de figuras y gráficos. Ejemplos de este tipo de análisis descriptivo pueden encontrarse en la prensa diaria, en la parte de información económico-social: series de tiempo, gráfica de barras, diagramas circulares, índices de precios, resultados de una encuesta y más elaborado, para más de una variable, en pirámide de edades, comparativas, etc. También puede encontrarse el uso de la estadística descriptiva en tablas de consumo, resultados deportivos, Accidentes laborales, Ventas anuales realizadas y, en general, hechos cuantificados en valores absolutos (tal cual), porcentajes (%) o índices (con un periodo base inicial = 100). En otras palabras, la Estadística Descriptiva se refiere a la recolección, presentación, descripción, análisis e interpretación de una colección de datos, esencialmente consiste en resumir éstos con uno o dos elementos de información (medidas descriptivas) que caracterizan la totalidad de los mismos. La estadística descriptiva es un

método de obtener de un conjunto de datos conclusiones sobre sí mismos y no sobrepasan el conocimiento proporcionado por éstos. Puede utilizarse para resumir o describir cualquier conjunto ya sea que se trate de una población o de una muestra, cuando en la etapa preliminar de la Estadística Inferencial se conocen los elementos de una muestra.

Suponga que un administrador calcula la producción promedio diaria de la empresa donde labora durante un mes. Como la estadística calculada describe el desempeño diario del departamento de producción en un mes determinado, pero no hace ninguna generalización acerca de los diferentes meses del año, podemos decir que el administrador está utilizando estadística descriptiva. Gráficas, tablas y diagramas que muestran los datos de manera que sea más fácil su entendimiento, son ejemplos de estadística descriptiva.

Supongamos que estamos laborando en una gran empresa, con un número muy elevado de trabajadores, por ejemplo 5,000, y queremos hacer un estudio estadístico sobre la altura de los trabajadores y las trabajadoras. Un método sería pasar departamento por departamento y medir a cada trabajador y trabajadora, esto nos podría llevar un tiempo considerable pero sería la forma más exacta de hacer dicho estudio, aunque es fácil encontrarnos con ausencias y tendríamos que volver varios días y pasar lista para conseguir la estatura de todo personal. Una vez que tengamos todos los datos en nuestro poder los resultados los obtendríamos mediante Estadística descriptiva. Otra posibilidad podría ser pasar departamento por departamento, y decirles a los

trabajadores y trabajadoras que anoten su estatura en un papel y después recoger la información. También así tendríamos un estudio de Estadística descriptiva, aunque seguramente menos fiable que con el método anterior, pues casi con toda seguridad, y lo digo por experiencia, algunos trabajadores y trabajadoras escribirán su estatura a cálculo y otros(as), con ganas de bromear, muy por encima o muy por debajo de la realidad. Y una última posibilidad sería escoger una muestra, es decir un grupo de por ejemplo 500 personas, hacer el estudio descriptivo sobre ellas y después generalizarlo a todo el personal con Estadística de inferencia. En este caso, comprobaríamos por una parte que cuanto mayor sea la muestra más trabajo tendremos, pero más fiable será el resultado final y por otra, que la elección de la muestra debe hacerse de manera que permita también fiarnos del resultado obtenido.

Supongamos que estamos laborando en una gran empresa, con un número muy elevado de trabajadores, por ejemplo 5,000, y queremos hacer un estudio estadístico sobre la altura de los trabajadores y las trabajadoras. Un método sería pasar departamento por departamento y medir a cada trabajador y trabajadora, esto nos podría llevar un tiempo considerable pero sería la forma más exacta de hacer dicho estudio, aunque es fácil encontrarnos con ausencias y tendríamos que volver varios días y pasar lista para conseguir la estatura de todo personal. Una vez que tengamos todos los datos en nuestro poder los resultados los obtendríamos mediante Estadística descriptiva. Otra posibilidad podría ser pasar departamento por departamento, y decirles a los trabajadores y trabajadoras que anoten su estatura en un papel y después recoger la información. También así tendríamos un estudio de Estadística descriptiva, aunque seguramente menos fiable

que con el método anterior, pues casi con toda seguridad, y lo digo por experiencia, algunos trabajadores y trabajadoras escribirán su estatura a cálculo y otros(as), con ganas de bromear, muy por encima o muy por debajo de la realidad. Y una última posibilidad sería escoger una muestra, es decir un grupo de por ejemplo 500 personas, hacer el estudio descriptivo sobre ellas y después generalizarlo a todo el personal con Estadística de inferencia. En este caso, comprobaríamos por una parte que cuanto mayor sea la muestra más trabajo tendremos, pero más fiable será el resultado final y por otra, que la elección de la muestra debe hacerse de manera que permita también fiarnos del resultado obtenido.

### ***La estadística descriptiva***

Es la rama de las matemáticas que recolecta, presenta o caracteriza un conjunto de datos por ejemplo: edad de una población, altura de los estudiantes de una escuela, temperatura en los meses de verano, etc. Esto con el fin de describir apropiadamente las diversas características de ese conjunto.

Realizando una revisión de la literatura podemos encontrar diversas fuentes que nos mencionan que la estadística es una herramienta que nos ayuda a conocer la realidad, sin embargo es importante considerar que también puede distorsionar la verdad, si no se tiene cuidado al utilizar los métodos estadísticos adecuados o si la interpretación se realiza de un manera errónea. La metodología estadística se emplea en muchos campos y áreas del conocimiento. Es

una disciplina que ayuda a diseñar el esquema de búsqueda y registro de información para describirla y analizarla con facilidad, mediante estimaciones, obtener conclusiones que enriquecen el conocimiento y describen la realidad.

Podemos observar que la estadística es una herramienta para mejorar la comprensión de los datos, pero sin perder de vista que el análisis de los datos deberá responder a los objetivos, hipótesis de la investigación.

Para su mejor comprensión y aplicación en el área de investigación la estadística presenta dos grandes divisiones estas son: descriptiva y la analítica.

Figura 1. Campos de aplicación de la estadística descriptiva



Para su mejor comprensión y aplicación en el área de investigación la estadística presenta dos grandes divisiones son la descriptiva y la analítica.

podemos observar que la estadística es una herramienta para mejorar la comprensión de los datos pero sin perder de vista que el análisis de los datos deberán responder a los objetivos, hipótesis, de la investigación.

para indagar en el conocimiento de la realidad concreta que al hombre le interesa, considera tres etapas fundamentales que son:

- Planear la búsqueda y la obtención de la información.
- sistematizar y organizar la información de tal forma que se pueda describir y analizar con facilidad.
- Efectuar inferencias sobre la realidad a partir de la información obtenida haciendo estimaciones o verificando hipótesis.

En cualquier estudio en el que los datos sean numéricos, el análisis de datos comienza con la estadística descriptiva también denominada estadística resumen. En la estadística descriptiva se pueden incluir distribuciones de frecuencia, las medidas de tendencia central, las medidas de dispersión y las puntuaciones estandarizadas. Cuando existe la posibilidad de manejar una gran cantidad de datos es importante considerar los paquetes estadísticos comerciales que fueron desarrollados para crear, editar, reportar y efectuar operaciones sobre una base de datos conservando la integridad y veracidad de ellos.

Los datos crudos obtenidos en una investigación provienen de un cuestionario, instrumento de recolección o mediciones realizadas a los sujetos de estudio; las respuestas recolectadas por si solas pueden no expresar nada, pero organizarlas de una manera adecuada y procesarlas represente más de lo que estamos buscando. Es necesario conformar una base de datos en donde se transformen las respuestas de los instrumentos o las mediciones en datos informativos y fáciles de analizar, es decir, que les asignemos códigos numéricos o de letras a las respuestas de acuerdo a la operacionalización de las variables, a este procedimiento se le denomina codificación de la información, siendo el objetivo tener una mayor facilidad en el manejo de los datos

Tabla 1. Codificación de variables

Variable	Categoría	Codificación	
		Números	Letras
<b>Sexo</b>	Masculino	Masculino= 1	Masculino= M
	femenino	Femenino= 2	Femenino= F
<b>Edad</b>	<10 años	<10 años=1	<10 años= X
	10-15 años	10-15 años= 2	10-15 años= Y
	16-20 años	16-20 años =3	16-20 años = Z

Fuente: Retomado a partir de Mendoza N. 2013

La construcción de la base de datos puede llevarse a cabo en una hoja cuadriculada o en una hoja de tabulación, cuando se realiza manualmente, o en la computadora. Existen tres formas para la recolección de datos desde una computadora. En

Directamente en un paquete estadístico: SPSS

Una vez que se han organizado los datos, ahora es importante describirlos, utilizando precisamente en tema en cuestión; estadística descriptiva.

Si las variables del estudio son nominales, la prueba estadística a elegir es una tasa, razón o proporción, siendo todas ellas medidas de frecuencia.

**Tasa:** es una relación en donde una frecuencia de sucesos se relaciona con la entidad más global, por lo que representa la frecuencia de un evento o atributo en relación a un grupo o población definida, multiplicada por una constante múltiplo de 10. Se coloca en el numerador el total de individuos con la característica estudiada y en el denominador el total de la población de la cual provienen.

**Razón:** es la expresión matemática de la relación entre entidades diferentes como numeradores y denominadores. Es el número de observaciones de un grupo con una característica dada dividido entre el número de observaciones, sin esa característica. Reflejan cambios relativos (actuales o potenciales) en dos cantidades, de manera simple una razón puede considerarse una comparación de tasas, en donde se llama razón de tasas.

**Proporción:** es el número de observaciones con una característica específica dividido entre el número total de observaciones de un grupo dado, es decir, es una parte dividida entre un todo.

Si la escala de medición de las variables es ordinal, las medidas de descripción estadística son la mediana y los cuartiles 25 y 75. Estas pruebas también se aplican en datos cuantitativos, ya sean discretos o continuos.

### ***Medidas de Tendencia Central***

**Media (Promedio):** es la medida de tendencia central que suele referirse como promedio. Se utiliza con mayor frecuencia, muchas de las pruebas importantes de significación estadística se basan en la media. Cuando los

investigadores trabajan con mediciones intercalares o de razón, la media casi siempre es la estadística que más se retoma.

**Moda:** es la más simple de las tres medidas para determinar la tendencia central, esta se determina mediante la inspección de la distribución de frecuencia

**Mediana:** es un punto en una escala numérica a partir del cual 50% de los casos están por encima y 50% por debajo de él. La mediana es un índice de la posición promedio en una distribución de números. Suele ser el índice preferido de tendencia central cuando la distribución es sesgada y cuando el interés del investigador radica en un valor típico, en mediciones de escala ordinal o superior.

Cuando una distribución de valores es simétrica y unimodal, coinciden los tres índices de tendencia central. En distribuciones con sesgo, difieren los valores de moda, mediana y media. Cuando la distribución es asimétrica, es preferible reportar los tres valores en lugar de seleccionar un índice único<sup>6</sup>.

### ***Medidas de Dispersión***

Las medidas de dispersión mayormente utilizadas son el rango, la varianza y la desviación estándar, con frecuencia se emplean cuadros de dispersión para ilustrar los datos. A continuación se describen cada una de ellas:

**Rango:** es la medida más simple de dispersión que se obtiene restando la puntuación más baja de la más alta. El rango es una puntuación diferencial, que emplea únicamente las dos puntuaciones extremas para establecer una comparación.

**Varianza:** se calcula mediante una ecuación matemática. El valor numérico obtenido a partir del cálculo depende de

la escala de medición empleada; el valor de la varianza calculado no tiene valor absoluto y se puede comparar únicamente con los datos obtenidos usando medidas similares.

**Desviación estándar:** es la raíz cuadrada de la varianza, igual que la media es el valor medio, la desviación estándar es el valor de la diferencia (desviación) media. Esta proporciona una medida de la desviación media de un valor a partir de la media en una muestra concreta. Indica el grado de error que se cometería si se usara solamente la media para interpretar los datos.

En la curva normal, el 68% del valor se encontrara dentro de una desviación estándar por encima o por debajo de la media, el 95% estará dentro de dos desviaciones estándar por encima o por debajo de la media y el 99% estará dentro de tres desviaciones estándar por encima o por debajo de la media.

**Cuartiles:** los percentiles o la centésima parte de la frecuencia total. Comúnmente los puntos de división utilizados con los llamados cuartiles (Q), en donde la frecuencia total se divide en 4 partes equivalentes al 25% cada una, de tal manera que se tienen tres cuartiles: Q25 o Q 1 al 25% de los datos Q50 o Q2 al 50% de los datos y Q75 o Q3 al 75% de los datos. Nótese que Q50 corresponde a la mediana. Como son posiciones no puede haber decimales, así que se aproximan al número inmediato superior, contando de menor a mayor en la base de datos ordenada.

**Proporción:** es el número de observaciones con una característica específica dividido entre el número total de

observaciones de un grupo dado, es decir, es una parte dividida entre un todo.

Si la escala de medición de las variables es ordinal, las medidas de descripción estadística son la mediana y los cuartiles 25 y 75. Estas pruebas también se aplican en datos cuantitativos, ya sean discretos o continuos.

### Medidas de Tendencia Central

- Media (Promedio): es la medida de tendencia central que suele referirse como promedio. Se utiliza con mayor frecuencia, muchas de las pruebas importantes de significación estadística se basan en la media. Cuando los investigadores trabajan con mediciones intervalares o de razón, la media casi siempre es la estadística que más se retoma.

- Moda: es la más simple de las tres medidas para determinar la tendencia central, esta se determina mediante la inspección de la distribución de frecuencia

- Mediana: es un punto en una escala numérica a partir del cual 50% de los casos están por encima y 50% por debajo de él. La mediana es un índice de la posición promedio en una distribución de números. Suele ser el índice preferido de tendencia central cuando la distribución es sesgada y cuando el interés del

investigador radica en un valor típico, en mediciones de escala ordinal o superior.

Cuando una distribución de valores es simétrica y unimodal, coinciden los tres índices de tendencia central. En distribuciones con sesgo, difieren los valores de moda, mediana y media. Cuando la distribución es asimétrica, es preferible reportar los tres valores en lugar de seleccionar un índice único.

### Medidas de Dispersión

Las medidas de dispersión mayormente utilizadas son el rango, la varianza y la desviación estándar, con frecuencia se emplean cuadros de dispersión para ilustrar los datos. A continuación se describen cada una de ellas:

**Rango:** es la medida más simple de dispersión que se obtiene restando la puntuación más baja de la más alta. El rango es una puntuación diferencial, que emplea únicamente las dos puntuaciones extremas para establecer una comparación.

**Varianza:** se calcula mediante una ecuación matemática. El valor numérico obtenido a partir del cálculo depende de la escala de medición empleada; el valor de la varianza calculado no tiene valor absoluto y se puede comparar

únicamente con los datos obtenidos usando medidas similares.

Desviación estándar: es la raíz cuadrada de la varianza, igual que la media es el valor medio, la desviación estándar es el valor de la diferencia (desviación) media. Esta proporciona una medida de la desviación media de un valor a partir de la media en una muestra concreta. Indica el grado de error que se cometería si se usara solamente la media para interpretar los datos.

En la curva normal, el 68% del valor se encontrara dentro de una desviación estándar por encima o por debajo de la media, el 95% estará dentro de dos desviaciones estándar por encima o por debajo de la media y el 99% estará dentro de tres desviaciones estándar por encima o por debajo de la media.

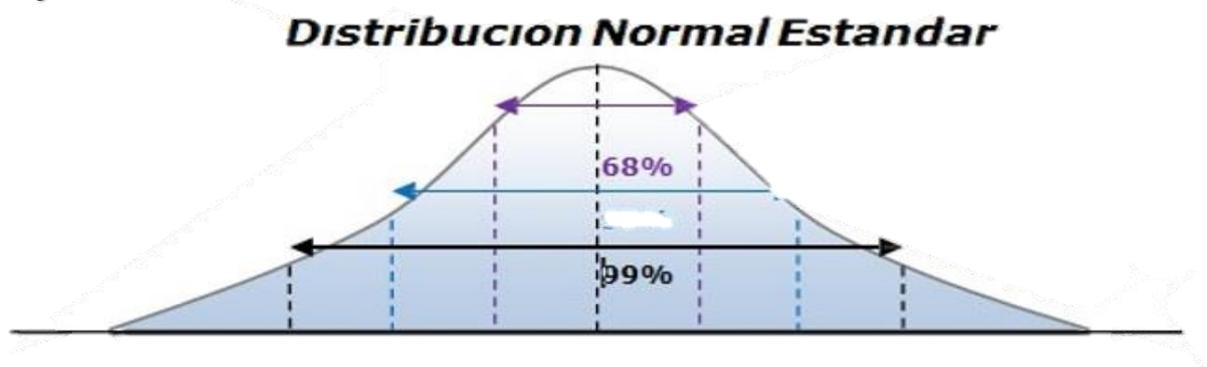


Figura 2. Distribución normal de la desviación estándar, tomada de Carroll N. 2017

## Tabla (de distribución ;de frecuencia para una ,dos o múltiples estradas )

A menudo en una investigación se recogen grandes cantidades de datos numéricos. Cuando esto ocurre es difícil visualizar un orden o estructura que ayude a analizarlos. Para lograrlo es necesario condensar los datos en grupos de acuerdo a ciertas divisiones de la recta numérica (intervalos o clases). Aunque con esta agrupación la información inicial sobre cada dato individual se pierde, es más fácil visualizar rápidamente las características principales del grupo total de datos

La frecuencia de un intervalo es el número de datos que corresponden a ese intervalo. Una distribución de frecuencia es una tabla en la que aparecen todos los intervalos y las frecuencias de datos correspondientes a cada intervalo. Esta agrupación de datos numéricos por intervalos o clases se llama una distribución de frecuencia porque en ella se indica cuan frecuentemente aparecen datos en cada intervalo.

Las tablas de distribución de frecuencias se utilizan cuando se recolectan datos, con ellas se pueden representar los datos de manera que es más fácil analizarlos.

Se pueden elaborar tablas de distribución de frecuencias para datos no agrupados y para datos agrupados. Estas últimas se utiliza cuando se tienen muchos datos.

Para elaborar tablas de distribuciones de frecuencia se debe tener en cuenta lo siguiente:

Cuando hay muchos datos se agrupan en clases.

Clase es cada uno de los grupos en que se dividen los datos. Para determinar cuántas clases crear, se puede utilizar la siguiente fórmula (fórmula de Sturges):

Número de clases =  $1 + 3,322 \log n$  donde n es el número total de datos.

Si al aplicar la fórmula se obtiene un número decimal, se aproxima al siguiente entero.

El intervalo de clase o el ancho de la clase (tamaño de la clase) es el espacio que hay entre el límite superior y el límite inferior de la clase, los cuales corresponden a los valores extremos de la clase. Para obtener el ancho de clase se utiliza la siguiente fórmula:

Ancho de clase =  $(\text{dato superior} - \text{dato inferior}) / \text{número de clases}$

La frecuencia absoluta es el número de veces que se repite cada dato. Cuando se agrupan los datos, es el número de datos que tiene cada clase. Se simboliza con  $f_j$ .

La marca de clase es el punto medio de la clase. Se obtiene dividiendo entre dos la suma de los valores extremos de cada clase.

El rango es la diferencia entre el valor mayor y el valor menor en estudio de una distribución de datos.

La frecuencia absoluta acumulada es la frecuencia total hasta el límite superior de cada clase. Se simboliza con  $F_j$ .

La frecuencia relativa de un dato da información sobre qué parte de la población o de la muestra en estudio corresponde a la característica analizada. Se obtiene dividiendo la frecuencia absoluta entre el número total de datos y se puede expresar como una fracción, como un decimal o como un porcentaje. Se simboliza con  $f_j / n$  donde  $n$  es el número de datos.

La frecuencia acumulada relativa es la frecuencia relativa total hasta el límite superior de cada clase. Se simboliza con  $F_j / n$  donde  $n$  es el número total de datos. Ejemplo:

Suponga que un investigador desea determinar cómo varía el peso de un grupo de estudiantes de primer semestre de una universidad. Selecciona una muestra de 50 estudiantes y registra sus pesos en kilogramos. Los datos obtenidos fueron los siguientes:

65	63	65	63	69	67	53	58	60	61
64	65	64	72	68	66	55	57	60	62
64	65	64	71	68	66	56	59	61	62
63	65	63	70	67	66	57	59	61	62
64	64	63	69	67	66	58	60	61	62

Para determinar el número de veces que aparece cada dato (frecuencia absoluta), se utiliza el

diagrama de tallo y hojas. Se traza una línea y a la izquierda se escriben las cifras anteriores a las unidades que tengan los datos, a la derecha de la línea se escriben la cifra de las unidades para cada uno de los datos. Este diagrama facilita determinar la cantidad de veces que se repite un dato y los valores de los datos con el fin de escribirlos de manera ordenada en la tabla.

<del>64</del>	<del>65</del>	<del>64</del>	<del>71</del>	<del>68</del>	<del>66</del>	<del>56</del>	<del>59</del>	<del>61</del>	<del>62</del>
<del>63</del>	<del>69</del>	<del>67</del>	<del>53</del>	<del>58</del>	<del>65</del>	<del>63</del>	<del>65</del>	<del>60</del>	<del>61</del>
<del>57</del>	<del>59</del>	<del>61</del>	<del>62</del>	<del>63</del>	<del>65</del>	<del>63</del>	<del>70</del>	<del>67</del>	<del>66</del>
<del>64</del>	<del>65</del>	<del>64</del>	<del>72</del>	<del>68</del>	<del>66</del>	<del>55</del>	<del>57</del>	<del>60</del>	<del>62</del>
<del>67</del>	<del>66</del>	<del>58</del>	<del>60</del>	<del>61</del>	<del>62</del>	<del>64</del>	<del>64</del>	<del>63</del>	<del>69</del>
5	6	9	3	8	7	9	5	7	8
6	4	5	4	8	6	1	2	3	9
	1	2	3	5	3	7	6	4	5
	6	0	1	2	4	4	3	9	
7	1	0	2						

Luego, se organiza la información en la tabla, de la siguiente manera:

### Tabla de datos no agrupados

	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Pesos (Kg)	$f_i$	$F_i$	$h_i$	$H_i$
53	1	1	2,00%	2,00%
55	1	2	2,00%	4,00%
56	1	3	2,00%	6,00%
57	2	5	4,00%	10,00%
58	2	7	4,00%	14,00%
59	2	9	4,00%	18,00%

Las gráficas que representan la información de la tabla son: **Histograma de frecuencias absolutas, Polígono de frecuencias, Diagrama circular, Ojiva.** Para construir la tabla de datos no agrupados se debe calcular primero lo siguiente:

**Número de clases o intervalos:**

$$k = 1 + 3,322 \log_{10} n$$

$$k = 1 + 3,322 \log_{10} 50 = 6,64 \approx 7$$

**Rango:**

$$\text{Rango} = (\text{dato máximo} - \text{dato mínimo})$$

$$\text{Rango} = 72 - 53 = 19$$

**Ancho de clase (tamaño del intervalo)**

$$I = \text{Rango} / k$$

$$I = 19 / 7 = 2,71 \approx 3$$

Al construir la tabla de datos agrupados con la información del ejemplo, se tiene:

1 **Tabla de datos agrupados**

	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Pesos (Kg)	$f_i$	$F_i$	$h_i$	$H_i$
53 – 55	2	2	4,00%	4,00%
56 – 58	5	7	10,00%	14,00%
59 – 61	9	16	18,00%	32,00%
62 – 64	15	31	30,00%	62,00%
65 – 67	12	43	24,00%	86,00%
68 – 70	5	48	10,00%	96,00%
71 – 73	2	50	4,00%	100,00%
	50		100,00%	

Para esta tabla también se pueden hacer histogramas o diagramas de barras y circulares.

Los estudios estadísticos que se centran en el análisis de una sola variable se llaman unidimensionales.

Sin embargo, en las situaciones reales es corriente que se tenga que investigar la combinación de dos variables estadísticas, en lo que se conoce por distribución bidimensional. Cuando se realizan estudios estadísticos de dos variables, el modelo resultante recibe el nombre de distribución bidimensional. Algunos ejemplos de este tipo de distribución son las tablas de altura-peso de un colectivo de población, la relación pulso-temperatura de un grupo de enfermos y la gráfica de ingresos y gastos de una empresa.

En estas distribuciones se manejan variables estadísticas bidimensionales, que constituyen pares de valores de cada una de las variables elementales que intervienen, denotados por  $(x_i, y_j)$ .

Para hacer un estudio de una variable estadística bidimensional, tendremos que recoger de cada individuo dos datos. Estos datos correspondientes a un mismo individuo no se pueden separar, pues vamos a estudiar la relación que existe entre ellos. Los podremos recoger formando pares de elementos  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , en los que el primer

elemento pertenece a la primera de las características observadas

y el segundo elemento a la segunda característica.

Una vez que hemos recogido todos los datos, la mejor forma de estudiarlos es disponerlos en una tabla estadística. Dependiendo de la cantidad de datos que tengamos, de la frecuencia de los mismos y de que para un mismo valor de una de las variables, se presenten varios valores de las otras, existen dos tipos de tablas estadísticas:

Una vez que hemos recogido todos los datos, la mejor forma de estudiarlos es disponerlos en una tabla estadística. Dependiendo de la cantidad de datos que tengamos, de la frecuencia de los mismos y de que para un mismo valor de una de las variables, se presenten varios valores de las otras, existen dos tipos de tablas estadísticas:

### **Tabla bidimensional simple**

Está formada por tres filas o columnas en las que se representan ordenadamente los valores de las variables y sus frecuencias. Está indicada para casos con pocos datos y pocos valores o ninguno repetidos

$X_i$	$X_1$	$X_2$	...	$X_i$	...	$X_m$	
$Y_i$	$Y_1$	$Y_2$	...	$Y_i$	...	$Y_n$	
$f_{ij}$	$f_{11}$	$f_{22}$	...	$f_{ij}$	...	$f_{mn}$	$N$

En caso de que las frecuencias sean iguales a uno, se puede omitir la fila o columna correspondiente a las mismas.

Escogiendo la fila o columna de cada una de las variables junto con la de las frecuencias, tenemos la tabla correspondiente a cada una de las variables unidimensionales, que podremos utilizar para calcular la media y la desviación típica de las mismas.

### Tabla de doble entrada

Está formada por tantas filas y columnas como valores tengamos de cada una de las variables, más una fila y una columna más para indicar los totales. Está indicada para casos con bastantes datos, en los que para cada valor de una variable, existen varios valores de la otra .

	$x_1$	$x_2$	...	$x_i$	...	$x_m$	Frecuencia absoluta de la variable Y
$y_1$	$f_{11}$	$f_{21}$	...	$f_{i1}$	...	$f_{m1}$	$\sum f_{i1}$
$y_2$	$f_{12}$	$f_{22}$	...	$f_{i2}$	...	$f_{m2}$	$\sum f_{i2}$
...	...	...	...	...	...	...	...
$y_i$	$f_{1i}$	$f_{2i}$	...	$f_{ii}$	...	$f_{mi}$	$\sum f_{ii}$
...	...	...	...	...	...	...	...
$y_n$	$f_{1n}$	$f_{2n}$	...	$f_{in}$	...	$f_{mn}$	$\sum f_{in}$
Frecuencia absoluta de la variable X	$\sum f_{1j}$	$\sum f_{2j}$	...	$\sum f_{ij}$	...	$\sum f_{jn}$	N

Escogiendo la primera y la última fila, tenemos la tabla estadística correspondiente a la primera variable unidimensional. Con la primera y última columnas construimos la tabla correspondiente a la segunda variable unidimensional. Estas dos distribuciones

reciben el nombre de distribuciones marginales. En la última celda aparecerá el total de la última fila y de la última columna, es decir, el número total de elementos estudiados (N).

Además, en esta tabla puede resultar de interés estudiar distribuciones unidimensionales correspondientes a un valor determinado de alguna de las variables, llamadas distribuciones condicionadas.

## Diagramas de caja

Un diagrama de cajas y bigotes es una manera conveniente de mostrar visualmente grupos de datos numéricos a través de sus cuartiles.

Las líneas que se extienden paralelas a las cajas se conocen como «bigotes», y se usan para indicar variabilidad fuera de los cuartiles superior e inferior. Los valores atípicos se representan a veces como puntos individuales que están en línea con los bigotes. Los diagramas de cajas y bigotes se pueden dibujar vertical u horizontalmente.

Normalmente utilizado en estadísticas descriptivas, los gráficos de cajas y bigotes son una excelente forma de examinar rápidamente uno o más conjuntos de datos gráficamente. Aunque parezcan primitivos en comparación con un Histograma o un Gráfico de Densidad, tienen la ventaja de ocupar menos espacio, lo cual es útil cuando se comparan distribuciones entre muchos grupos o conjuntos de datos.

Aquí están los tipos de observaciones que uno puede hacer al ver un diagrama de cajas y bigotes:

Cuáles son los valores clave, tales como: el promedio, el percentil 25 medio, etc.

Si hay valores atípicos y cuáles son sus valores.

Si los datos son simétricos.

Cuán estrechamente se agrupan los datos.

Si los datos están sesgados y si es así, en qué dirección.

Dos de las variaciones más comúnmente utilizadas de los diagramas de cajas y bigotes son: los diagramas de caja de anchura variable y los diagramas de caja con muescas.

Ejemplo 2: El tratamiento de los niños con desórdenes de la conducta puede ser complejo. El tratamiento se puede proveer en una variedad de escenarios dependiendo de la severidad de los comportamientos. Además del reto que ofrece el tratamiento, se encuentran la falta de cooperación del niño/niña y el miedo y la falta de confianza de los adultos. Para poder diseñar un plan integral

de tratamiento, el siquiatra de niños y adolescentes puede utilizar la información del niño, la familia, los profesores y de otros especialistas médicos para entender las causas del desorden. Para ello, un siquiatra local ha considerado una muestra aleatoria de 20 niños, anotando el tiempo necesario que requiere en cada niño para lograr un plan integral del tratamiento, obteniéndose lo siguiente (en horas): 6 7 7 8 8 8 8 9

9 9 9 9 9 10 10 10 10 10 11 a. Calcule las medidas de tendencia central y de dispersión de estos datos, indicando a qué tipo de medida pertenece. Respuesta: Medidas de tendencia

central:  $\frac{6+7+7+8+8+8+8+9+9+9+9+9+10+10+10+10+10+11}{20} = 8,8$   
 de la Mediana: Datos ordenados:

1° 2° 3° 4° 5° 6° 7° 8° 9° 10° 11° 12° 13° 14° 15° 16° 17° 18° 19°  
 20° 6 7 7 8 8 8 8 9 9 9 9 9 9 10 10

10 10 10 11  $Q1=8$   $Q2=9$   $Q3=10$  Posición de la Mediana  $(\frac{5+11}{2})=8$ , por tanto la mediana será el valor medio entre la décima y la undécima observación. Mediana = 9 horas. Moda = 9 horas (el valor que más se repite). Medidas de dispersión: Desviación estándar:  $s = 1,24$  horas. Rango =  $11 - 6 = 5$  horas. Cálculos para el rango entre cuartiles: El cuartil 1 será la mediana de los primeros 10 datos, es decir, se encuentra entre la quinta y sexta observación: Cuartil 1 = 8 horas. El cuartil 3 será la mediana de los últimos 10 datos, es decir, se encuentra entre la 15ava y 16ava observación: Cuartil 3 = 10 horas. Rango entre cuartiles =  $10 - 8 = 2$  horas.

Dibuje un diagrama de caja. Comente el resultado acerca de la distribución.

Respuesta: Para dibujar el gráfico de caja necesitamos verificar si existen valores extremos: Valores extremos:  $X_i < Q1 - 1,5 (Q3 - Q1)$   
 $6? 8 - 1,5 (10 - 8) = 8 - 3 = 5$  no es menor que 5, por lo tanto 6 no es un valor extremo.  $X_i > Q3 + 1,5 (Q3 - Q1)$   
 $11? 10 + 3 = 13$  no es mayor que 13, por lo tanto 11 no es un valor extremo. Distribución del tiempo necesario para diseñar un plan integral de un tratamiento que requiere un niño con desordenes de la conducta.

La caja muestra cierta simetría, aunque los bigotes dicen lo contrario, mostrando un sesgo a la izquierda. Dibuje un diagrama de tallo y hoja. Comente el resultado acerca de la distribución.

Respuesta: Distribución del tiempo necesario para diseñar un plan integral de un tratamiento que requiere un niño con desordenes de la conducta.

Stem & Leaf	1.00	6. 0	2.00	7. 00
4.00	8. 0000	7.00	9. 0000000	5.00
10. 00000	1.00	11. 0	Stan width:	1.00

Each leaf: 1 case(s) La distribución no es simétrica con un leve sesgo a la izquierda.

## Conclusión

Con todo lo que ya se ha investigado en este trabajo ,se puede concluir que la estadística descriptiva es una rama de las matemáticas que no se encuentra muy visible en lo cotidiano pero que en realidad es de mucha utilidad para interpretar y ver desde un punto de vista muy general datos que se obtienen.Cabe mencionar que dentro de la ingeniería civil la probabilidad juega un rol muy importante ya nos sirve por que esta basada en la recopilación de datos la cual estas son representadas mediante graficas estadísticos .En la cual se presentan ,los principales indicadores de la actividad constructora: producción, despachos

,ventas ,exportación e importación de cemento  
,venta interna de asfalto y la producción y venta de  
barras de construcción.

Otro beneficio que obtendríamos de esta materia es  
que nos permite realizar estudios reales, con  
poblaciones exactas; lo cual no ayuda a mejorar los  
proyectos de un negocio y así lograr mayor  
resultados logrando nuestros objetivos.

## Bibliografía:

<http://soystaff.blogspot.com/2015/10/aplicacion-de-la-estadística.html?m=1>

<http://unam.wordpress.com/2016/04/27/estadística-descriptiva/>

<http://datavizcatague.com/ES/metodos/diagrama-cajasybigotes.html>

<https://www.academia.edu/31925407/aplicación-de-la-estadística-en-el-proceso-administrativo-en-los-estudios-estadísticos>

de

.