



**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA**

**194 - Metodología de la Investigación**

Dra. Nelly Jeannette Alcántara Galdámez

**Tema:**

Cap.10 Análisis de datos cuantitativos

- Estadística Descriptiva
- Medidas de tendencia central
- Medias de Variabilidad

**PRESENTADO POR:**

20551123 Carlo Marcello Menjivar Montes de Oca

**CAMPUS SAN PEDRO SULA**

**07 de febrero de 2022**

*“El planteamiento del problema es el centro, el corazón de la investigación:  
dicta o define los métodos y la ruta a seguir.”*

- Roberto Hernández-Sampieri<sup>1</sup>

## Tabla de Contenido

Introducción.....	4
Objetivos .....	5
Marco Teórico .....	7
Conclusión.....	13
Bibliografía .....	14
Anexos.....	15

## Introducción

Este informe tiene como primera intención exponer, explicar y desarrollar los temas de medidas de tendencia central y medidas de variabilidad de la Estadística Descriptiva. Ambos temas se centran en los métodos para la interpretación de los resultados de los datos recolectados de una investigación; esto es necesario para poder analizarlos de forma cuantitativa.

Actualmente existen diferentes programas para analizar los datos recopilados en una investigación, el funcionamiento de todos es más o menos similar y, generalmente, incluyen las dos partes o segmentos que se mencionaron en el capítulo 9: una parte de definiciones de las **variables**, que a su vez explican los **datos** (los elementos de la codificación item por item o indicador por indicador), y la otra parte, la matriz de datos. *Para fines de este reporte hemos elegido utilizar un Cuaderno Jupyter, herramienta comúnmente utilizada mucho en la rama de ciencia de datos.*

Hemos optado por usar el texto *“Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta”*<sup>2</sup>, y seguiremos la normas Apa 2021<sup>3</sup>.

La segunda intención en este reporte es practicar el conocimiento adquirido en la clase, mezclar parte del conocimiento adquirido de otras clases que estoy cursando actualmente y, me disculpo por las alertas de plagio, admito he copiado y pegado bastante.

*“Agradezco su tiempo y el tiempo de mis compañeros así que espero disfruten del choque de media noche de música, maestría y unas tazas de café”*<sup>4</sup>.

## **Objetivos**

- Exponer y explicar el concepto de estadística descriptiva y su relación en el desarrollo de investigaciones con orientación cuantitativa.
- Definir y profundizar en los temas de medidas de tendencia central y medidas de variabilidad.
- Presentar el concepto, formula y ejemplos de los subtemas: media, mediana y moda para el tema de medidas de tendencia central.
- Presentar el concepto, formula y ejemplos de los subtemas: rango, desviación estándar y varianza.

*“Al analizar los datos cuantitativos debemos recordar dos cuestiones: primero, que los modelos estadísticos son representaciones de la realidad, no la realidad misma; y segundo, los resultados numéricos siempre se interpretan en contexto, por ejemplo, un mismo valor de presión arterial no es igual en un bebe que en una persona de la tercera edad.”*

– Roberto Hernández-Sampieri<sup>2</sup>

## Marco Teórico

El Capítulo 10 del libro, “Metodología de la investigación: Las rutas cuantitativas cualitativa”<sup>5</sup>, presenta el proceso general para realizar análisis estadístico en la ruta cuantitativa. En este capítulo se presenta el proceso general para realizar análisis estadístico en la ruta cuantitativa, etapa por etapa. Asimismo, se comentan y ejemplifican los métodos para establecer la confiabilidad, los elementos de la estadística descriptiva y las pruebas de la estadística inferencial, tanto paramétricas como no paramétricas y la forma de interpretar sus resultados, considerando que los análisis se efectúan en un programa computacional. Por ello, el capítulo se centra en los usos y la interpretación de los métodos, más que en los procedimientos de cálculo.

En este reporte solo se desarrollarán los temas de medidas de tendencia central y de medidas de variabilidad.

A continuación, definiremos el concepto de Estadística Descriptiva según la primera búsqueda de Google<sup>6</sup> (aunque muy acertada respuesta recomendando no confiar en todo lo que se le pregunta a Google sin antes, en palabras de Sampieri, entendamos muy bien el contexto).

### Estadística descriptiva

La estadística descriptiva es una disciplina que se encarga de recoger, almacenar, ordenar, realizar tablas o gráficos y calcular parámetros básicos sobre el conjunto de datos.

**Estadística descriptiva** (esta vez dejemos a Sampieri explicarlo antes que le dé un colapso por culpa de Google)

El investigador busca, en primer término, describir sus datos y posteriormente efectuar análisis estadísticos para relacionar sus variables. Es decir, realiza análisis de estadística descriptiva para cada una de las variables de la matriz (ítems o indicadores) y luego para cada una de las variables del estudio, finalmente aplica cálculos estadísticos para probar sus hipótesis. Los tipos o métodos

de análisis cuantitativo o estadístico son variados y se comentaran en el capítulo; pero cabe señalar que el análisis no es indiscriminado, sino que cada método tiene su razón de ser y un propósito específico; por ello, no debes hacer más análisis de los necesarios. La estadística no es un fin en sí misma, sino una herramienta para evaluar tus datos.

Perfecto. Ahora más o menos, casi que, entiendo lo que es la estadística descriptiva. ¿y ahora?

Bueno, la primer tarea es que describas los datos, valores o puntuaciones obtenidas para cada variable de la investigación como resultado de aplicar tu instrumento a la muestra o casos. Pero ¿cómo pueden describirse estos datos? Esto se logra al describir la distribución de las puntuaciones o frecuencias de tus variables.

### **¿Qué es una distribución de frecuencias?**

Una *distribución de frecuencias* es un conjunto de puntuaciones respecto de una variable ordenadas en sus respectivas categorías y generalmente se presenta como una tabla.

Existen diferentes maneras para presentar las distribuciones de frecuencias, especialmente cuando utilizas los porcentajes, y pueden presentar en forma de histogramas o gráficas de otro tipo. Este reporte solo presentará las medidas de tendencia central.

### **¿Cuáles son las medidas de tendencia central?**

Las medidas de tendencia central son los valores medios de la distribución de frecuencias y son útiles para ubicar los resultados de la muestra en la escala de medición de la variable en cuestión. Las medidas de tendencia central principales son tres: moda, mediana y media. El nivel de medición de la variable establece cuales de ellas son apropiadas para interpretar.

La **moda** o **modo** es la categoría, puntuación o intervalo (o punto medio de este) que ocurre con



mayor frecuencia. Se utiliza con todos los niveles de medición. Imagina que realizas un estudio sobre relaciones sentimentales en tu universidad y en el cuestionario preguntas si tienen o no pareja y obtienes los siguientes resultados (tabla 10.10). La moda sería “2” o “no tienen pareja”.

La **mediana** es el valor que divide a la distribución o escala presentada exactamente por la mitad. Esto es, el 50% de los casos caen por debajo de la mediana y el restante 50% por encima. La mediana refleja la posición intermedia de la distribución de frecuencias. Por ejemplo, si en tu salón la mediana de edad es de 20 años, esto significa que la mitad sobrepasa esta edad y el otro medio es más joven. La mediana mundial es de 28.4 años (Central Intelligence Agency, 2015). Constituye una medida de tendencia central propia de los niveles de medición ordinal, por intervalos y de razón. No tiene sentido con variables nominales, porque en este nivel no hay jerarquías ni noción de “encima o debajo”.

La **media** es la medida de tendencia central más utilizada y puede definirse como el promedio aritmético de una distribución. Se simboliza como  $\bar{X}$  y es la suma de todos los valores dividida entre el número de casos. Es una medida que únicamente se aplica a mediciones por intervalos o de razón, pues carece de sentido para variables medidas en un nivel nominal u ordinal. Es sensible a los valores extremos.

Por ejemplo, hasta junio de 2017 la media de goles por partido de Lionel Messi era de 0.81 y la de Cristiano Ronaldo de 0.70 (incluyendo clubes y selecciones) (Acosta, 2017). Estos valores son altos o bajos dependiendo de la distribución. Por ejemplo, un promedio de calificación de 5 en una materia es bajo si la escala va de 0 a 10, pero no si esta oscila entre 0 y 5.

### ¿Cuáles son las medidas de variabilidad?

Las **medidas de la variabilidad** indican la dispersión de los datos en la escala de medición de la variable en cuestión y responden a la pregunta sobre donde están diseminadas las puntuaciones obtenidas para la muestra o casos. Las medidas de tendencia central son valores en una distribución y las medidas de la variabilidad son intervalos que designan distancias o un número

de unidades en la escala que utilizaste para medir la variable. Las medidas de la variabilidad más utilizadas son rango, desviación estándar y varianza.

El rango, también denominado recorrido, es la diferencia entre la puntuación superior y la puntuación inferior, e indica el número de unidades en la escala de medición que se necesitan para incluir los valores máximo y mínimo. Por ejemplo, si en una muestra la persona mayor tiene 30 años y la menor 17, el rango será: 13. Cuanto más grande sea el rango, mayor será la dispersión de los datos de una distribución.

La desviación estándar o típica es el promedio de desviación de las puntuaciones de la muestra respecto a la media. Esta medida se expresa en las unidades originales de medición de la distribución. Cuanto mayor sea la dispersión de los datos alrededor de la media, mayor será la desviación estándar. Se simboliza con: s, DE, Desv. Est. (MinitabR), Desviación tip. (SPSSR) o la letra sigma minúscula  $\sigma$ .

Supongamos que un investigador obtuvo para su muestra una media (promedio) de ingreso familiar mensual de 10 000 pesos (o la moneda de tu país) y una desviación estándar de 1 000 pesos. La interpretación es que los ingresos familiares de la muestra se desvían, en promedio, mil unidades monetarias respecto a la media.

La desviación típica solo se utiliza en variables medidas por intervalos o de razón.

La **varianza** es la desviación estándar elevada al cuadrado y se simboliza  $s^2$ . Es un concepto estadístico muy importante, que fundamenta muchas pruebas cuantitativas. Diversos métodos estadísticos parten de la descomposición de la varianza. Sin embargo, con fines descriptivos se utiliza preferentemente la desviación estándar.

### **¿Como se interpretan las medidas de tendencia central y de la variabilidad?**

Los programas de análisis te proporcionan las medidas de tendencia central y de dispersión para

cada una de las variables de tu estudio, y al describir los resultados, las interpretas en conjunto, no aisladamente. Consideras todos los valores, veamos un ejemplo.

Supón que la aplicas a 248 personas. SPSSR, MinitabR y otros programas te proporcionaran un resultado de salida (output) más o menos así (con variaciones en los diversos programas o versiones, pero siempre busca la información pertinente).

La actitud hacia el presidente es “regular”. 50% está por encima de 3 y la restante mitad se ubica por debajo de este valor. El promedio es 3.177. En cambio, si los resultados fueran:

Moda: 1 Mediana: 1.5 Media ( $\bar{X}$ ): 1.3 Desviación estándar: 0.4  
Máximo: 3.0 Mínimo: 1.0 Rango: 2.0

La interpretación sería que la actitud hacia el presidente es muy desfavorable. En la figura 10.13 vemos gráficamente la comparación de resultados entre un resultado de una actitud favorable y uno de actitud muy desfavorable en la muestra. La variabilidad también es menor en el caso de esta última (los datos se encuentran menos dispersos).

### **¿Hay alguna otra estadística descriptiva?**

Si, la asimetría (*skewness* en inglés) y la curtosis. Los polígonos de frecuencia son curvas, por ello se representan como tales (figura 10.14), para que puedan analizarse en términos de probabilidad y visualizar su grado de dispersión. Estos dos elementos resultan esenciales para analizar estas curvas o polígonos de frecuencias.

La asimetría es una estadística necesaria para conocer cuánto se parece nuestra distribución a una distribución teórica llamada curva normal (la cual se representa también en la figura 10.14) y constituye un indicador del lado de la curva donde se agrupan las frecuencias. Si es cero (asimetría = 0), la curva o distribución es simétrica. Cuando es positiva, quiere decir que hay más valores agrupados hacia la izquierda de la curva (por debajo de la media). Cuando es negativa,

significa que los valores tienden a agruparse hacia la derecha de la curva (por encima de la media).

La curtosis es un indicador de lo plana o “picuda” que es una curva. Cuando es cero (curtosis = 0), significa que puede tratarse de una curva normal. Si es positiva, quiere decir que la curva, la distribución o el polígono es mas “picudo” o elevado. Si la curtosis es negativa, indica que es más plana la curva.

La asimetría y la curtosis requieren al menos un nivel de medición por intervalos. En la figura 10.14 se muestran ejemplos de curvas con su interpretación<sup>7</sup>.

Debes recordar que en una investigación obtienes una distribución de frecuencias y las estadísticas descriptivas para cada variable, las que se necesiten de acuerdo con los propósitos de la investigación y los niveles de medición. Los programas también pueden agruparte los resultados principales de todas tus variables si así lo solicitas.

## **Conclusión**

Debes recordar que en una investigación obtienes una distribución de frecuencias y las estadísticas descriptivas para cada variable, las que se necesiten de acuerdo con los propósitos de la investigación y los niveles de medición. Los programas también pueden agruparte los resultados principales de todas tus variables si así lo solicitas.

En este reporte se expusieron y explicaron los conceptos de estadística descriptiva y su relación en el desarrollo de investigaciones con orientación cuantitativa.

Se profundizó en la explicación en los temas de medidas de tendencia central y medidas de variabilidad.

Presentaron los conceptos, formulas y ejemplos de los subtemas: media, mediana y moda para el tema de medidas de tendencia central. Así mismo para los subtemas: rango, desviación estándar y varianza.

## Bibliografía

**1** de acuerdo a Roberto Hernández-Sampieri (2018, p. 38) respecto a la primera edición...

*Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta.*

**2** Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación:*

*Las rutas cuantitativa, cualitativa y mixta.*

**3** Citar con Normas APA 2021, codeinep.

<https://codeinep.org/normas-apa-2021/>

**4** estoy practicando las normas APA así que el entrenamiento va incluido.

**5** Hernández Sampieri, R., & Mendoza Torres, C. P. (2018). *Metodología de la investigación:*

*Las rutas cuantitativa, cualitativa y mixta.*

**6** literal, pueden Googlear si no me creen... qué complicada forma de explicar las cosas tiene Google.

**9** Citar con Normas APA 2021, codeinep.

<https://codeinep.org/normas-apa-2021/>

## Anexos

### 7 Figura 10.14. Ejemplos de curvas o distribuciones y su interpretación.



Figura 10.14. Ejemplos de curvas o distribuciones y su interpretación.

### 8 Figura 10.1: Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta

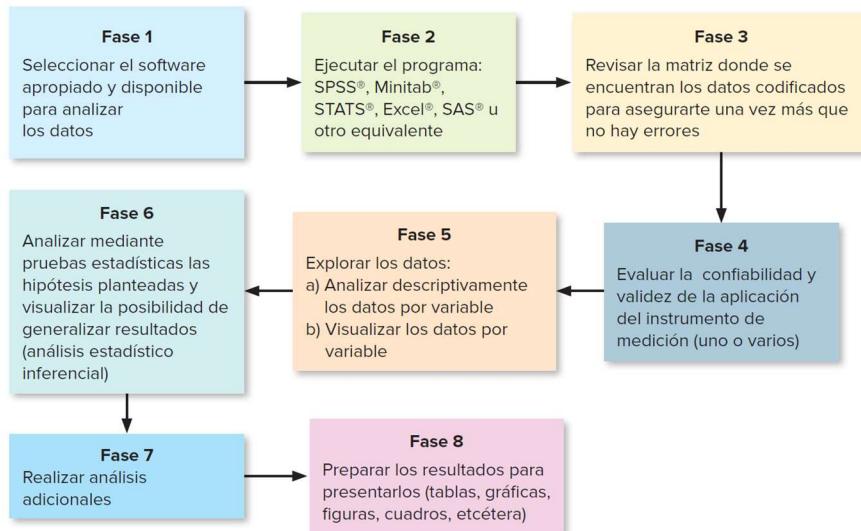


Figura 10.1. Proceso para efectuar análisis estadístico en un programa computacional.

## **Procesos para describir los valores del ejercicio**

### **Fase o paso 1:**

#### **seleccionar el programa de análisis o software**

Actualmente tienes disponibles diferentes programas para analizar tus datos integrados en la matriz de codificación. El funcionamiento de todos es muy similar y, generalmente, incluyen las dos partes o segmentos que se mencionaron en el capítulo anterior: una parte de definiciones de las variables, que a su vez explican los datos (los elementos de la codificación ítem por ítem o indicador por indicador), y la otra parte, la matriz de datos. La primera parte es para que se comprenda la segunda. Las definiciones, desde luego, las preparas tu como investigador. Lo que haces, una vez recolectados tus datos, es precisar los parámetros de la matriz de datos en el programa (nombre de cada variable en la matriz —que equivale a un ítem, reactivo, indicador, categoría o subcategoría de contenido u observación—, tipo de variable o ítem, ancho en dígitos, etc.) e introducir o capturar estos datos en la matriz, la cual es como cualquier hoja de cálculo. Asimismo, recuerda que la matriz de datos tiene columnas (variables unidimensionales, ítems o indicadores), filas o renglones (casos o unidades de muestreo o análisis) y celdas (intersecciones entre una columna y un renglón). Cada celda contiene un dato (que significa un valor de un caso en una variable unidimensional o ítem).

### **Fase 2: ejecutar el programa**

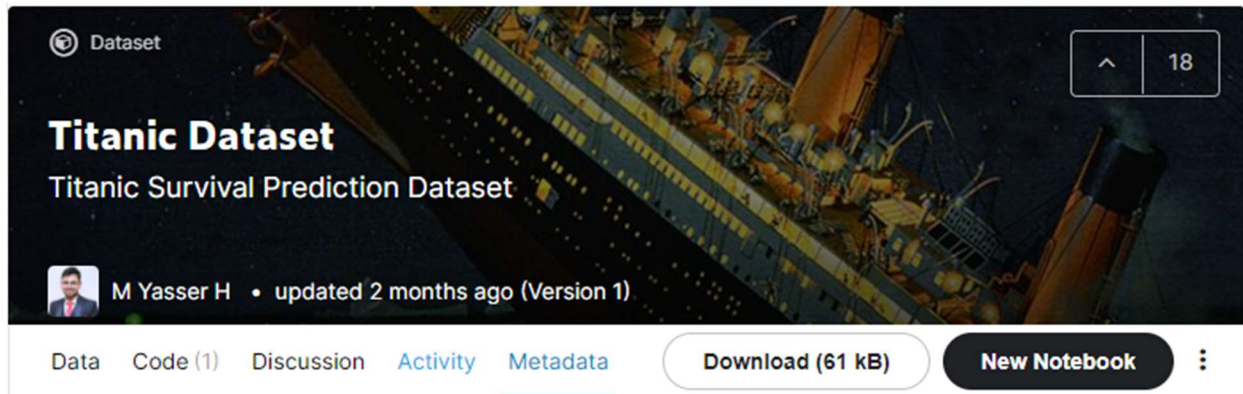
La mayoría de los programas son fáciles de usar, pues lo único que tienes que hacer es solicitar los análisis requeridos seleccionando las opciones apropiadas.

### **Fase 3: revisión de la matriz**

En todos los programas abres la matriz creada por ti con la finalidad de verificar que no existan errores de captura (visualmente y con la instrucción respectiva para revisión de la matriz). Además, puedes solicitar la distribución de frecuencias de las variables (casos en las categorías)



del estudio para ver si en alguna variable hay errores de codificación (o bien, de las columnas o indicadores, ítems o equivalentes).



Dataset

# Titanic Dataset

Titanic Survival Prediction Dataset

M Yasser H • updated 2 months ago (Version 1)

Data Code (1) Discussion Activity Metadata Download (61 kB) New Notebook

<https://www.kaggle.com/yasserh/titanic-dataset>



Dataset

# Height of Male and Female by Country 2022

Average Height of Male and Female by Country

Majyhain • updated 21 days ago

Data Code (4) Discussion (1) Activity Metadata Download (8 kB) New Notebook

<https://www.kaggle.com/majyhain/height-of-male-and-female-by-country-2022>