

COLECCIÓN

INVESTIGACIÓN, DESARROLLO
E INNOVACIÓN

4



ESTADÍSTICA DESCRIPTIVA CON UN ENFOQUE DE DESARROLLO LOCAL SOSTENIBLE

RAÚL LÓPEZ FERNÁNDEZ
ARTURO BOFILL PLACERES
DIANA ELISA PALMERO URQUIZA
MARÍA DEL CARMEN FRANCO GÓMEZ

UMET
UNIVERSIDAD
METROPOLITANA

ESTADÍSTICA DESCRIPTIVA

CON UN ENFOQUE

DE DESARROLLO LOCAL SOSTENIBLE

RAÚL LÓPEZ FERNÁNDEZ
ARTURO BOFILL PLACERES
DIANA ELISA PALMERO URQUIZA
MARÍA DEL CARMEN FRANCO GÓMEZ

INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Con el auspicio de la Fundación Metropolitana



ESTADÍSTICA DESCRIPTIVA CON UN ENFOQUE DE DESARROLLO LOCAL SOSTENIBLE

**RAÚL LÓPEZ FERNÁNDEZ
ARTURO BOFILL PLACERES
DIANA ELISA PALMERO URQUIZA
MARÍA DEL CARMEN FRANCO GÓMEZ**

Diseño de carátula: D. I. Yunisley Bruno Díaz

Composición de textos: D. I. Yunisley Bruno Díaz

Corrección: MSc. Dolores Pérez Dueñas

Dirección editorial: Dr. C. Jorge Luis León González

Sobre la presente edición:

© Editorial Universo Sur, 2019

© Universidad Metropolitana de Ecuador, 2019

ISBN: 978-959-257-549-3

Podrá reproducirse, de forma parcial o total, siempre que se haga de forma literal y se mencione la fuente.



Editorial: “Universo Sur”.

Universidad de Cienfuegos. Carretera a Rodas, Km 3 ½.

Cuatro Caminos. Cienfuegos. Cuba.

CP: 59430

E-mail: eus@ucf.edu.cu

Prólogo

El desarrollo de la economía y la sociedad en este Siglo XXI está marcado por la globalización, el desarrollo de la informática y las comunicaciones, la acumulación de grandes volúmenes de información, su procesamiento y la rapidez con que ocurren los cambios, todo ello regido, fundamentalmente, por fenómenos de carácter aleatorio y cualquier estudio para la toma de decisiones conlleva indispensablemente la utilización de las Probabilidades y la Estadística. De aquí que en la mayoría de las carreras universitarias se incluyan estas disciplinas con carácter obligatorio.

La Estadística Descriptiva, también hoy identificada como estudio exploratorio de datos, es una de las ramas de la Estadística y su objetivo fundamental es convertir los datos en información para la toma de decisiones. El desarrollo de la informática ha permitido procesar grandes volúmenes de datos en información en breves segundos y se han desarrollado paquetes informáticos que contienen las herramientas de la Estadística Descriptiva. Otras técnicas que hoy se mencionan frecuentemente para el procesamiento de datos como la Minería de Datos y el Big Data utilizan frecuentemente esta herramienta en sus procedimientos.

En el campo de la gestión empresarial la Estadística Descriptiva tiene una amplia utilización para la toma de decisiones. Baste solo mencionar su uso en la gestión de la calidad, la gestión financiera, de mantenimiento, de logística y de estudios de mercados.

Todo lo anterior ha conllevado a modificar la forma de la enseñanza de la Estadística Descriptiva, haciendo más énfasis en que se conozca los elementos esenciales de cada una de sus componentes, su procesamiento usando la informática y sobre todo la interpretación de sus resultados.

El libro que se presenta ha seguido el enfoque mencionado anteriormente. Presenta brevemente la fundamentación teórica de las distintas técnicas de la Estadística Descriptiva, utiliza paquetes informáticos para el procesamiento de los datos, en este caso la Hoja de Cálculo Excel y el SPSS y enfatiza la interpretación de los resultados. La selección de los paquetes informáticos se debe a la sencillez y generalización de las Hojas de Cálculo, que hoy pueden llevarse en un teléfono inteligente y el SPSS por ser uno de los paquetes más utilizados en el campo de la investigación científica. A diferencia de otros textos sobre la materia, este desarrolla ejemplos, problemas resueltos y propuestos, asociados a situaciones prácticas y reales que prepara adecuadamente al lector para la aprehensión de esta materia.

El texto está diseñado para su utilización en los programas de pregrado de Estadística Descriptiva que se imparten en las carreras universitarias fundamentalmente asociadas a la gestión empresarial y a la toma de decisión bajo riesgo. Puede también ser utilizado en cursos de postgrado que tengan en su programa esta asignatura.

Los contenidos tratados son: Definiciones de Estadística. Conceptos fundamentales: tipos de variables. Pasos en un estudio estadístico. Distribución de Frecuencias: absoluta, acumulada y relativa. Representación gráfica de las frecuencias. Medidas de tendencia central. Medidas de dispersión o variabilidad. Medidas de posición. Diagrama de caja.

Los autores agradecen la colaboración de otros colegas en la versión final que hoy se presenta, quienes con oportunas sugerencias han ayudado a mejorar su contenido. También han sido importantes las opiniones de los estudiantes que han utilizado parte del texto en las clases que se han impartido por el colectivo de profesores. Toda obra humana es posible de ser mejorada, por lo que agradecemos cualquier opinión o sugerencia que pueda enriquecer esta primera edición.

Introducción

Cuando se habla de Estadística, se suele pensar en una relación de datos numéricos presentada de forma ordenada. Esta idea, consecuencia del concepto popular que existe sobre el término se extiende cada vez más debido a la influencia del entorno en la Sociedad Moderna, ya que hoy en día es casi imposible dejar de percibir en cualquier medio de difusión, periódico, radio, televisión, etc, los más disímiles tipos de información estadística.

Solo cuando nos adentramos de manera más específica en el campo de la investigación en las más diversas ramas del saber, ya sea en el campo de las Ciencias Sociales, la Medicina, la Biología, etc, empezamos a percibir a la Estadística no como una simple vía para presentar la información, sino que se convierte en una **“herramienta”** indispensable para el trabajo en cualquier profesión, que nos permite comprender mejor la esencia de los fenómenos y procesos con los que acostumbramos a relacionarnos y de tal manera obtener resultados y por tanto beneficios, en cualquier tipo de estudio, cuyo comportamiento, por su variabilidad intrínseca, no puede ser abordado desde la perspectiva de las leyes deterministas.

Se puede, desde un punto de vista más amplio, definir la estadística como: la Ciencia que se ocupa de cómo reunir, organizar y analizar datos numéricos y que ayuda a resolver problemas a través de las actividades investigativas y la toma de decisiones.

Otra definición desde un punto de vista más detallado, puede darnos visión algo más concreta desde la perspectiva de los contenidos de nuestro texto, veamos:

La Estadística es la Ciencia que se ocupa de: la sistematización, recogida, ordenación y presentación de los datos referentes a un fenómeno o proceso que presenta variabilidad o incertidumbre para su estudio metódico, con el objetivo de deducir las leyes que rigen los mismos y poder de esa manera hacer predicciones sobre esos fenómenos o procesos, tomar decisiones y establecer conclusiones.

A partir de ella se destaca las diferentes partes componentes de la Estadística como Ciencia y que serán objeto de estudio en los diferentes temas planteados: Como es el primero de una serie de tres libros, se ha decidido presentar, en los primeros momentos, elementos generales de la Estadística para su mejor comprensión.

Se debe tener en cuenta que las fuentes de las tablas “datos ficticios”, es con el objetivo de conservar la ética científica. El Excel utilizado es la versión 16, la cual tiene incorporado las facilidades gráficas del histograma y el diagrama de caja. En esta obra se ha tratado de utilizar en cada ejercicio todas las posibles soluciones, lo cual da un enfoque sistémico a la metodología utilizada.

Estadística Descriptiva

Es la parte de la Estadística que se ocupa de la sistematización, recogida, ordenación y presentación de los datos referentes a un fenómeno o proceso que presenta variabilidad o incertidumbre para su estudio metódico.

Probabilidad

Es la parte de la Estadística que se ocupa de deducir las leyes que rigen los fenómenos o procesos que presentan variabilidad o incertidumbre para su estudio metódico.

Estadística inferencial

Es la parte de la Estadística que se ocupa de hacer predicciones sobre esos fenómenos o procesos, tomar decisiones y establecer conclusiones.

Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o cosas. Hacia el año 3000 a.C. los babilonios usaban pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y sobre los géneros vendidos o cambiados mediante trueque.

En el siglo XXXI a.C., mucho antes de construir las pirámides, los egipcios analizaban los datos de la población y la renta del país. Los libros bíblicos de Números y Crónicas incluyen, en algunas partes, trabajos de estadística. El primero contiene dos censos de la población de Israel y el segundo describe el bienestar material de las diversas tribus judías. En China existían registros numéricos similares con anterioridad al año 2000 a.C. Los griegos clásicos realizaban censos cuya información se utilizaba hacia el 594 a.C. para cobrar impuestos.

El Imperio romano fue el primer gobierno que recopiló una gran cantidad de datos sobre la población, superficie y renta de todos los territorios bajo su control. Durante la edad media solo se realizaron algunos censos exhaustivos en Europa. Los reyes Pipino el Breve y Carlomagno ordenaron

hacer estudios minuciosos de las propiedades de la Iglesia en los años 758 y 762 respectivamente.

Después de la conquista normanda de Inglaterra en 1066, el rey Guillermo I de Inglaterra encargó la realización de un censo. La información obtenida con este censo, llevado a cabo en 1086, se recoge en el Domesday Book.

El registro de nacimientos y defunciones comenzó en Inglaterra a principios del siglo XVI, y en 1662 apareció el primer estudio estadístico notable de población, titulado *Observations on the London Bills of Mortality* (Comentarios sobre las partidas de defunción en Londres). Un estudio similar sobre la tasa de mortalidad en la ciudad de Breslau, en Alemania, realizado en 1691, fue utilizado por el astrónomo inglés Edmund Halley como base para la primera tabla de mortalidad.

En el siglo XIX, con la generalización del método científico para estudiar todos los fenómenos de las ciencias naturales y sociales, los investigadores aceptaron la necesidad de reducir la información a valores numéricos para evitar la ambigüedad de las descripciones verbales.

En nuestros días, la estadística se ha convertido en un método efectivo para describir con exactitud los valores de datos económicos, políticos, sociales, psicológicos, biológicos o físicos, y sirve como herramienta para relacionar y analizar dichos datos.

El trabajo del experto estadístico no consiste ya solo en reunir y tabular los datos, sino sobre todo en el proceso de “interpretación” de esa información. El desarrollo de la teoría de la probabilidad ha aumentado el alcance de las aplicaciones de la estadística. Muchos conjuntos de datos se pueden aproximar, con gran exactitud, utilizando determinadas distribuciones probabilísticas; los resultados de éstas se pueden utilizar para analizar datos estadísticos. La probabilidad es útil para comprobar la fiabilidad de las inferencias estadísticas y para predecir el tipo y la cantidad de datos necesarios en un determinado estudio estadístico.

Capítulo I. Conceptos fundamentales en estadística

1.1. Diferentes tipos de variables estadísticas

A continuación, se relacionan algunos conceptos básicos indispensables para el trabajo con la Estadística y que serán de uso común a todo lo largo del curso.

Individuos o elementos; son aquellas personas, seres vivos (plantas o animales) u objetos que poseen una determinada información que deseamos estudiar.

Población; es el conjunto de individuos o elementos sobre el que estamos interesados en realizar un determinado estudio y que generalmente tiene la característica de ser demasiado grande para poder abarcarla en su totalidad.

Muestra; es un subconjunto de la población que deseamos estudiar, al que objetivamente tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones), para lo cual se requiere que sea representativa de la población y que por lo tanto estará conformado por individuos elementos seleccionados de la misma.

Variables; son aquellas características observables (medibles) que varían entre los individuos o elementos de la población y que constituyen el centro de nuestro objeto de estudio en la misma (López, et al., 2017).

Cuando hablemos de **variable** haremos referencia a un símbolo (X, Y, A, B, ...) que puede tomar cualquier **modalidad** (valor) de un conjunto determinado, que llamaremos **dominio de la variable o rango**. En función del tipo de dominio, o sea del carácter ó modalidad que la misma pueda asumir, las variables las clasificamos como **categorías y numéricas**. En muchos libros de texto a la variable categórica se le denomina cualitativa y a la numérica, cuantitativa (Aragón, 2016).

Son variables categóricas, aquellas que sus valores (*modalidades*) no se pueden asociar naturalmente a un número (no se pueden hacer operaciones algebraicas con ellos). Son ejemplo de este tipo de variables, el sexo (M y F), la nacionalidad, la categoría ocupacional, época del año, nivel de ataque de

una plaga a una cosecha y nivel de satisfacción de un cliente.

Las variables **categorías o cualitativas**, pueden a su vez diferenciarse entre ellas de acuerdo a las posibilidades de las mismas para ser ordenadas y pueden clasificarse como:

Nominales: Sus valores no se pueden ordenar, por ejemplo: sexo, categoría ocupacional, nacionalidad, defectos de un producto.

Ordinales: Sus valores se pueden ordenar; por ejemplo: grado de satisfacción de un cliente (muy satisfecho; satisfecho; medianamente satisfecho; insatisfecho), grado de limpieza de las habitaciones en un hotel.

Las variables **numéricas o cuantitativas** tienen valores numéricos y se pueden realizar operaciones algebraicas con ellas. Este tipo de variables se clasifica en dos tipos:

Discretas: Si toma valores enteros, es decir, siempre que se pueda hacer una biyección del conjunto de los números naturales con el espacio muestral objeto de estudio, por ejemplo: Número de camiones en una empresa de carga, número de habitaciones de un hotel, número de trabajadores en una empresa.

Continuas: Si entre dos valores, son posibles infinitos valores intermedios, es decir, siempre que no se pueda hacer una biyección del conjunto de los números naturales con el espacio muestral objeto de estudio, por ejemplo: altura de una persona, velocidad de un auto, peso de un producto (López, et al., 2017).

1.2. Pasos en un estudio estadístico

En este epígrafe, se describe someramente los diferentes pasos de los que debe contar un estudio estadístico. En esta descripción aparecerán conceptos generales muchos de los cuales aún no han sido objeto de estudio, por lo que no es necesario comprenderlo todo aún. Sin embargo, esta representación ayudará a esclarecer algunos de los conceptos ya descritos y servirá de guía para los distintos temas que serán tratados en este texto.

1. Plantear el objetivo de un estudio sobre una población.

- Ejemplo: Caracterizar la calidad del servicio en un hotel.
- ¿Qué tipo de categoría de hotel? ¿Todos los servicios o solo algunos?

2. Decidir qué datos recoger. (*Diseño*)

- Qué individuos pertenecerán al estudio (*Muestras*)
- Criterios de exclusión, exclusión y salida, ¿Cómo se eligen?
- Qué datos recoger de los mismos (*Variables*)

3. Recoger los datos. (*Muestreo*)

4. Resumir los datos obtenidos. (*Estadística Descriptiva*)

- Elaborar tablas de frecuencias, gráficas.
- Calcular estadígrafos de tendencia central y dispersión.

5. Realizar una inferencia sobre la población y cuantificar la confianza en la inferencia. (*Inferencia Estadística*)

- Realizar comparaciones, pruebas de hipótesis.
- Estudios de relaciones entre las variables estudiadas.

1.3. Resumir los datos de un estudio estadístico. Distribución de frecuencias

Una de las primeras acciones que se acometen al realizar un estudio estadístico y que compone la esencia de lo que denominamos Estadística Descriptiva, lo constituye la organización y resumen de los datos. Una de las herramientas esenciales en este paso es la Distribución Frecuencial. Esta técnica permite que, a partir de un grupo de datos, los mismos se agrupen en lo que se denomina clases o intervalos y para ello se tiene en cuenta el tipo de variable que se está trabajando

La primera de estas denominaciones (Clases), se utiliza en la organización de los datos de las variables categóricas o cualitativas fundamentalmente y la segunda (Intervalos) generalmente para las variables numéricas o cuantitativas.

Esta agrupación de los datos en clases o intervalos, permite conocer la Distribución Frecuencial Absoluta, esto es la cantidad de datos que se ubica en cada clase o intervalo. Se expondrán algunos ejemplos de cómo determinar la Distribución Frecuencial Absoluta para ambos tipos de variables.

EJEMPLO 1. En una Agencia de Viajes se ofertan 3 tipos de paquetes turísticos: Sol y Playa (SP), Ecoturismo (ET) y Turismo de aventura (TA). Se ha seleccionado una muestra de 70 paquetes vendidos en los meses anteriores, lo que se muestra en la Tabla 1.

Tabla 1. Datos ficticios de una Agencia de Viajes. Ecuador, 2018.

SP	TA	SP	SP	ET	ET	SP	SP	TA	ET
SP	TA	ET	ET	ET	TA	SP	SP	ET	TA
SP	ET	SP	SP	SP	TA	ET	ET	SP	SP
SP	ET	SP	ET	SP	ET	SP	TA	ET	SP
SP	SP	ET	TA	SP	SP	ET	SP	TA	ET
ET	TA	SP	SP	SP	TA	ET	ET	SP	ET
ET	SP	SP	ET	TA	SP	ET	SP	SP	ET

Fuente: Datos ficticios

La variable de este problema es de tipo categórica o cualitativa y tiene tres clases o categorías: Sol y Playa (SP); Ecoturismo (ET) y Turismo de Aventuras (TA)

Realizando un conteo de los datos de la muestra que se ubican en cada clase se puede obtener la siguiente Tabla de Frecuencia Absoluta:

Sol y Playa (SP) ///// ///// ///// ///// ///// ///// // 33

Ecoturismo (ET) ///// ///// ///// ///// ///// 25

Turismo de aventura (TA) ///// ///// // 12

CLASE	CANTIDAD
Sol y Playa (SP)	33
Ecoturismo (ET)	25
Turismo de aventura (TA)	12
TOTAL	70

Esta forma de agrupación permite tener una información sobre las ventas de la Agencia de Viajes, donde la mayor cantidad se concentra en el turismo de Sol y Playa y en el Ecoturismo.

Este caso se trata de una variable de modalidad cualitativa **nominal** (Levine & Berenson, 2006). En muchos problemas de aplicación de la Estadística a la gestión, las variables categóricas son codificadas con números enteros para poder aplicar algunas herramientas. En este caso la codificación podría ser: Sol y Playa (1); Ecoturismo (2); Turismo de Aventura (3). Veamos otro caso de este tipo de variable; pero **ordinal**.

EJEMPLO 2. A través de una encuesta se obtiene el grado de satisfacción por los servicios prestados en un hotel, estableciéndose para ello las siguientes clases: Muy satisfecho, satisfecho, Moderadamente satisfecho e Insatisfecho; aplicadas las encuestas a un total de 20 clientes los resultados fueron.

Tabla 2. Satisfacción por los servicios prestados en un hotel (X). Ecuador, 2018

CLASES	CANTIDAD
Muy satisfecho	4
Satisfecho	7
Moderadamente satisfecho	5
Insatisfecho	4
TOTAL	20

Fuente: Datos ficticios.

La Tabla 2, muestra directamente la Frecuencia Absoluta para ese estudio.

Para la agrupación de datos, cuando estos son numéricos o cuantitativos, las clases donde se clasifica cada dato se identifica por medio de **intervalos numéricos**. Para ilustrar este caso se desarrolla el siguiente ejemplo:

EJEMPLO 3. Se quiere estudiar la duración del servicio de un cliente que accede a la caja de un Banco. Para ello se han tomado 50 observaciones de esa variable y el valor de las mismas se muestran en la Tabla 3.

Tabla 3. Duración del servicio de un cliente que accede a la caja de un Banco. Ecuador, 2018.

OBS No.	TIEMPO (minutos)								
1	18	11	13	21	24	31	15	41	16
2	12	12	16	22	14	32	21	42	10
3	15	13	15	23	19	33	18	43	21
4	19	14	21	24	21	34	22	44	20
5	21	15	18	25	15	35	15	45	16
6	24	16	23	26	24	36	17	46	19

7	13	17	20	27	16	37	23	47	11
8	19	18	14	28	18	38	20	48	16
9	16	19	15	29	22	39	18	49	18
10	26	20	18	30	19	40	15	50	21

Fuente: Datos ficticios.

Como puede apreciarse en este registro primario, la información así recogida no ofrece una visión tan clara como en los casos anteriores en los que la posibilidad de hacerlo por clases lo facilita; por lo que resulta necesario un ordenamiento por **intervalos**.

Para construir una Tabla de Frecuencia Absoluta por intervalo, se requiere definir los intervalos. En la construcción de los intervalos, se tendrán en cuenta determinadas condiciones prácticas, por lo que será necesario, que el investigador la adapte a las condiciones reales deseadas. Los pasos (Santoyo, 2001), a realizar son los siguientes:

1. Identificar el menor y el mayor valor de los datos.

En el ejemplo de la tabla 3, el menor será el valor de 10 minutos, que se corresponde con la observación número 42 y el mayor valor es 26 que se corresponde con la observación número 10.

2. Calcular el rango de los datos

Rango = Valor Mayor – Valor menor (conocido también como recorrido)

Rango = 26 – 10 = 16 minutos

3. Fijar el número de intervalos deseados para la agrupación.

Este aspecto es decisión del que realiza el estudio. Generalmente el número de intervalos puede estar entre 5 y 15 pero puede ser mayor si el investigador así lo desea. Habrá que considerar que los datos no queden ni muy concentrados ni muy dispersos y que en cada uno de los intervalos exista algún dato u observación.

4. Determinar la amplitud de cada intervalo.

La amplitud de cada intervalo, se determina como:

Amplitud del intervalo = Rango / Número de intervalos

Si se decidiera agrupar los datos anteriores en 6 intervalos, entonces la amplitud de estos intervalos será:

$$\text{Amplitud } 16 / 6 = 2.66 \text{ minutos}$$

Para fines prácticos esta amplitud calculada puede aproximarse a un valor cercano superior. En este ejemplo como los datos son números enteros, se escogió la amplitud como 3 minutos.

Otro método es dividir por 5 y por 20 el recorrido, (5 y 20 se escoge porque el investigador necesita el resumen de los datos en al menos 5 clases y a lo sumo 20 clases) en el intervalo resultante escoger un número. Los autores de este libro proponen que el número escogido sea el mayor entero impar.

$$\text{Ejemplo: } 16/5 = 3,2$$

$$16/20 = 0,8$$

Intervalo resultante: [0,8; 3,2]

El mayor número entero del intervalo es 3.

5. Calcular los límites de cada intervalo.

Los límites de cada intervalo, son determinados por el investigador, considerando el número de intervalo, ya definido y considerando los siguientes aspectos:

- Todas las observaciones deben quedar ubicadas en alguno de los intervalos.
- Debe quedar claramente establecido los límites de cada intervalo de forma tal que cada observación se ubique en uno y solo un intervalo. Hay diversas formas de cumplir con esta condición.

Para iniciar el cálculo de los límites de los intervalos se debe asignar un valor inicial para el límite inferior del primer intervalo que debe ser el menor valor o un valor inferior cercano a ese número. Cada intervalo tendrá la amplitud seleccionada como aproximación a la calculada.

En el ejemplo, para 6 intervalos, se asume como amplitud el valor de 3 minutos y podrían construirse los siguientes intervalos:

[9 – 12]

[10 – 13]

[12 – 15]

[13 – 16]

[15 – 18]

[16 – 19]

[18 – 21]

[19 – 22]

[21 – 24]

[22 – 25]

[24 – 27]

[25 – 28]

Ambos conjuntos cumplen con las condiciones que agrupan a todas las observaciones y que cada observación se ubica en un único intervalo. En este texto se selecciona el número más pequeño, en el ejemplo 10, para el límite inferior del primer intervalo y a partir de este se adiciona el valor de la amplitud para encontrar el resto de los intervalos.

¿Qué pasa cuando un número coincide con el límite superior de un intervalo y el inferior del siguiente intervalo? En el ejemplo el valor de 13 minutos podría ubicarse en el primero o en el segundo intervalo. Para evitar este problema, se debe asumir que siempre que se presente esa situación, la observación se contará en el intervalo que coincida con el límite superior. Así en el ejemplo el valor de 13 se colocaría en el primer intervalo, el de 16 en el segundo y así sucesivamente. También puede asumirse lo contrario, pero lo que se asuma se mantiene en todo el proceso de ubicación de las observaciones en un intervalo.

Con los intervalos determinados puede construirse la Tabla de Frecuencia Absoluta (Tabla 4). Utilizando el primer conjunto de intervalos, se tiene:

Tabla 4. Distribución de frecuencia absoluta de la variable “Duración del servicio de un cliente”. Ecuador, 2018.

INTERVALO (minutos)	CONTEO OBSERVACIONES	FRECUENCIA
10 - 13	/////	5
13 – 16	///// ///// /////	15
16 – 19	///// ///// ///	13
19 – 22	///// ///// /	11
22 – 25	/////	5
25 - 28	/	1
TOTAL		50

Fuente: Datos ficticio.

Hasta el momento la agrupación de datos en Tablas de Frecuencias, solo ha considerado la cantidad de observaciones en cada clase o intervalo y que se denomina **frecuencia absoluta**. Esta frecuencia se denota como f_i , donde i es la identificación de la clase o intervalo

Existen otros tipos de frecuencia, que ayudan al análisis más completo de la información obtenida con los datos agrupados, estas son:

Frecuencia relativa, de la clase i es el cociente que resulta de dividir la frecuencia absoluta de dicha clase y el número total de observaciones, es decir

$$f_{ri} = \frac{f_i}{n}$$

Donde:

f_{ri} : Frecuencia relativa de la clase i .

f_i : Frecuencia absoluta de la clase i .

n : Total de datos u observaciones. (Spiegel & Stephens, 2009).

Obsérvese que f_{ri} es el *tanto por uno* de observaciones que están en la clase i . Multiplicado por 100 representa el porcentaje de la población que comprende esa clase.

Frecuencia absoluta acumulada F_k , es la suma de las frecuencias hasta un determinado intervalo k y permite conocer el número de observaciones que es igual o menor que el límite superior de ese intervalo.

$$F_k = n_1 + n_2 + \dots + n_k = \sum_{i=1}^k n_i$$

Donde:

n_i : Número de observaciones en la clase o intervalo $i = 1, 2, 3 \dots, m$

Frecuencia relativa acumulada, F_{rk} es la suma de las frecuencias relativas hasta el intervalo k , siendo el tanto por uno de los elementos de la población que están con un valor menor o igual al límite superior del intervalo k . (Thomas, 2014).

$$F_{rk} = f_{r1} + f_{r2} + \dots + f_{rk} = \sum_{i=1}^k f_{ri}$$

Donde:

F_{rk} : Frecuencia relativa acumulada hasta el intervalo k.

f_{ri} : Frecuencia relativa del intervalo para $i = 1, 2, \dots, m$

De forma general se denomina **Distribución de Frecuencias** al conjunto de clases junto a las frecuencias correspondientes a cada una de ellas.

Las **Tablas de Frecuencias** que sirven para presentar de forma ordenada las distribuciones de frecuencias, se describen a partir de las consideraciones apuntadas anteriormente, o sea en función de la determinación de acuerdo a las clases o intervalos, calculando las frecuencias Absolutas, Relativas, Absolutas Acumuladas y Relativas acumuladas.

De forma general su composición se muestra en la Tabla 5

Tabla 5. Distribución de frecuencia de la variable “para cualquier variable”. Ecuador, 2018.

Clase ó Intervalo	Frec. Abs.	Frec. Abs. Acum.	Frec. Rel.	Frec. Rel. Acumu.
i	f_i	F_i	f_{ri}	F_{ri}
1	f_1	F_1	f_{r1}	F_{r1}
2	f_2	F_2	f_{r2}	F_{r2}

m	f_m	$F_m = n$	f_{rm}	$F_{rm} = 1$
TOTAL	n		1	

Fuente: Elaboración propia.

EJEMPLO 4. Aplicando lo anterior al ejemplo 3 de la duración del servicio en la caja de un banco, se puede construir la Tabla 6:

Tabla 6. Distribución de frecuencia de la variable “Duración del servicio de un cliente”: Ecuador, 2018.

Intervalo	Frecuencia Absoluta.	Frecuencia Absoluta Acumulada	Frecuencia Relativa Absoluta	Frecuencia Relativa Acumulada
10 - 13	5	5	0.10	0.10
13 – 16	15	20	0.30	0.40
16 – 19	13	33	0.26	0.66
19 – 22	11	44	0.22	0.88
22 – 25	5	49	0.10	0.98
25 - 28	1	50	0.02	1.00
	50		1	

Fuente: Datos ficticios.

Esta forma de ordenamiento de los datos nos ofrece una visión mucho más completa acerca de la estructura de los mismos y resulta de gran ayuda para el análisis y la toma de decisiones. De la Tabla anterior puede concluirse lo siguiente:

- Ya que la frecuencia relativa para el intervalo entre 13 y 16 minutos es de 0.30, se puede concluir que el 30% de los clientes tienen un tiempo de servicio entre 13 y 16 minutos
- La frecuencia relativa de los dos intervalos (13 -16 y 16 - 19) suma 0.56. Esto implica que más de la mitad de los clientes que arriban a la caja (56%) demoran entre 13 y 19 minutos.
- En la columna de la frecuencia relativa acumulada, puede observarse, que solo un 10% de los clientes demoran 13 minutos o menos en la atención de la caja y por tanto un 0.90 ($1 - 0.10$), o sea el 90% demoran más de 13 minutos.
- Si uno de los objetivos del banco es que un cliente que llegue a la caja demore menos de 12 minutos, puede concluirse que este objetivo se está incumpliendo y será necesario realizar un análisis más detallado para la toma de decisiones sobre este objetivo.

EJEMPLO 5. En el caso del problema de la Satisfacción de los clientes, en el que se trata de una variable cualitativa, que puede ser considerada como ordinal, se tendrá:

Tabla 7. Estudio satisfacción de clientes. Ecuador, 2018.

Clases	Frecuencia
Muy satisfecho	4
Satisfecho	7
Moderadamente satisfecho	5
Insatisfecho	4

Fuente: Datos ficticios.

Para este caso se obtiene la tabla 8, de la distribución de frecuencias:

Tabla 8. Distribución de frecuencia de la “satisfacción del cliente” Ecuador, 2018.

Clases	Frecuencia Absoluta	Frecuencia Absoluta Acumulada	Frecuencia Relativa Absoluta	Frecuencia Relativa Acumulada
Muy satisfecho	4	4	0.200	0.200
Satisfecho	7	11	0.350	0.550
Moderadamente satisfecho	5	16	0.250	0.800
Insatisfecho	4	20	0.200	1.000
Total	20		1.0	

Fuente: Datos ficticios.

De la tabla 8 se puede concluirse que:

- La cifra más alta de encuestados (frecuencia absoluta de 7) manifiesta estar satisfecho con los servicios que se brindan en el hotel.
- La frecuencia acumulada entre las categorías de muy satisfecho y satisfecho es de 0.55, lo que implica que más de la mitad (55%) de los clientes tiene buena opinión de los servicios que brinda el hotel.
- Lo anterior implica que el resto, 45%, es moderadamente satisfecho o insatisfecho, lo que resulta un valor alto, más si se tiene en cuenta que un 20%

está insatisfecho. Esto debe implicar una revisión por la Administración del Hotel de las causas que conllevan a estas evaluaciones.

1.4. Aplicación de la informática al cálculo de frecuencias

El desarrollo de la informática ha sido un elemento fundamental en el uso de las herramientas estadísticas, al elaborarse paquetes informáticos que permiten procesar los datos u observaciones de un experimento en forma rápida y exacta. Esto ha permitido un amplio uso de técnicas que por su complejidad en el procesamiento de los datos no eran utilizadas frecuentemente, como los diseños experimentales, modelos de regresión y modelos de múltiples variables. En la actualidad existen un grupo de paquetes informáticos de amplio uso, tales como el SPSS, STATGRAPHIC Y MINITAB, por mencionar algunos de los más utilizados. También las Hojas de Cálculo EXCEL trae un amplio grupo de funciones estadísticas, sobre todo de las utilizadas en la Estadística Descriptiva y las Probabilidades. Considerando que al EXCEL se tiene acceso de forma sencilla, utilizando incluso los teléfonos celulares, en este Capítulo se hará uso del mismo para apoyar el procesamiento de datos en algunos temas (Rico, 2016).

Para el caso del cálculo de la Frecuencia Absoluta, existen dos funciones en EXCEL que permiten su cálculo considerando si la variable quiere organizarse en clases o en intervalos. Como se ha visto se utiliza la organización frecuencial en clases, cuando las variables son cualitativas y en intervalo cuando son cuantitativas. En el caso de variables cuantitativas discretas, estas pueden organizarse de ambas formas.

Para el caso de querer determinar la Frecuencia Absoluta para una variable cualitativa, utilizando el Excel, se seleccionará la función "CONTAR SI".

Para ejemplificar su uso, se utilizará los datos del Ejemplo 1.1 de la Agencia de Viajes.

El primer paso es colocar los datos en una hoja Excel, agregando en una columna debajo de la tabla los valores que toma la variable y en la columna inmediata a la derecha el título FREC ABSOLUTA tal como se muestra en la Figura 1.

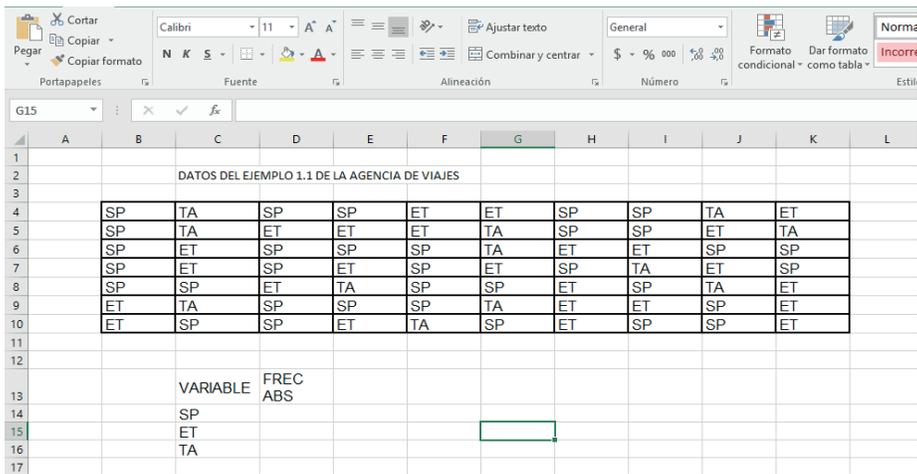


Figura 1. Ventana inicial del Excel para Tabla de Frecuencia Absoluta.

Para calcular la Frecuencia Absoluta, haga clic en la casilla asociada a la primera variable (D14) y busque en las funciones estadísticas de Excel la función CONTAR.SI, tal como se muestra en la Figura 2.

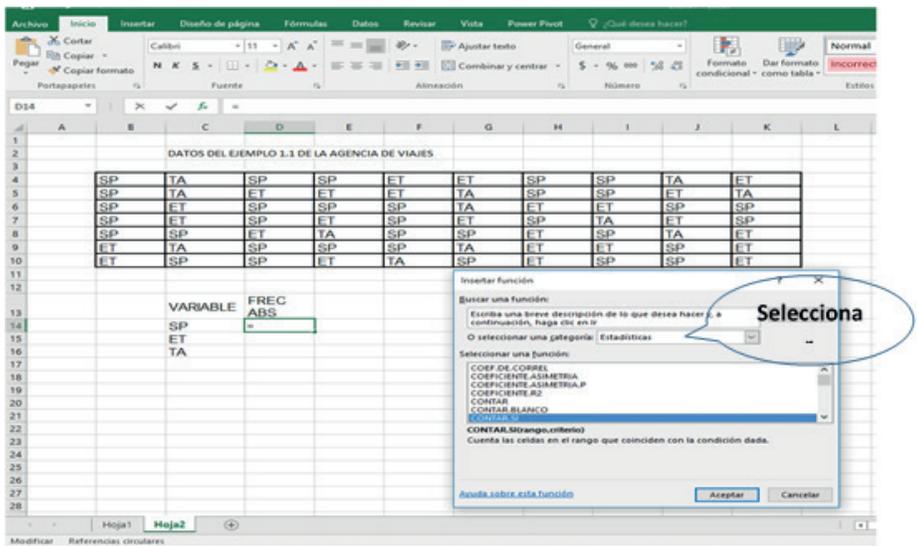


Figura 2. Función CONTAR.SI en Excel.

Marque aceptar y aparecerá el cuadro de diálogo que se muestra en la Figura 3:

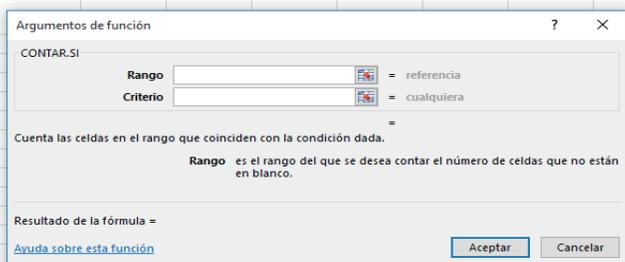


Figura 3. Cuadro de diálogo.

En la ventana, donde dice “Rango” coloque la matriz de datos y en “Criterio” las casillas de la columna donde aparecen las variables, tal como se muestra en la Figura 4.

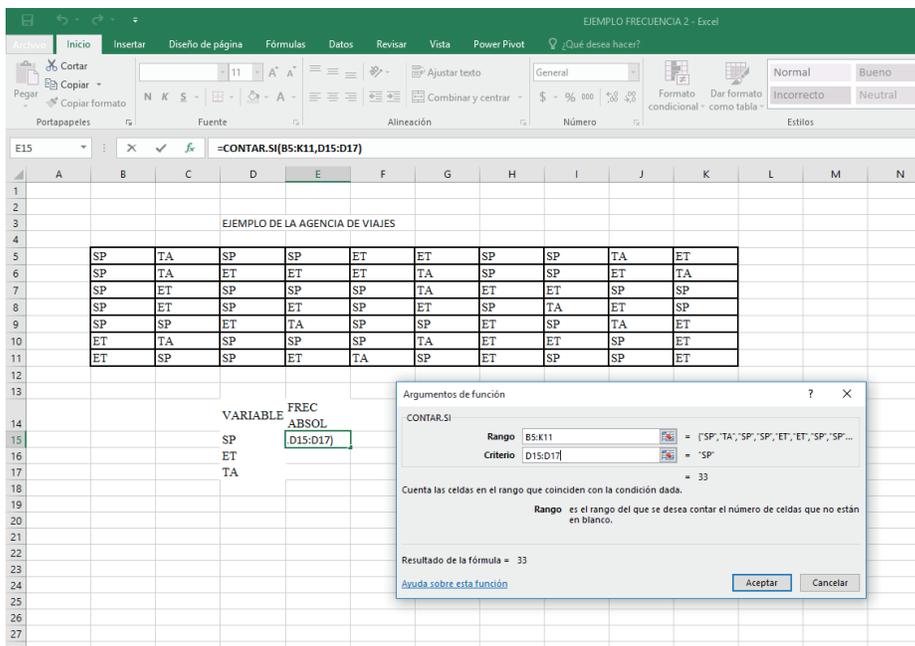


Figura 4. Argumento de la función “CONTAR.SI”.

Marque aceptar y aparecerá la Frecuencia Absoluta asociada a la primera variable (SP). Para encontrar la Frecuencia del resto de las variables, marque las casillas asociadas a la Frecuencia Absoluta y de a las teclas F2 y después

CONTROL –MAYUSCULA – ENTER y aparecerá el resto de las Frecuencias Absolutas, tal como se muestra en la Figura 5.

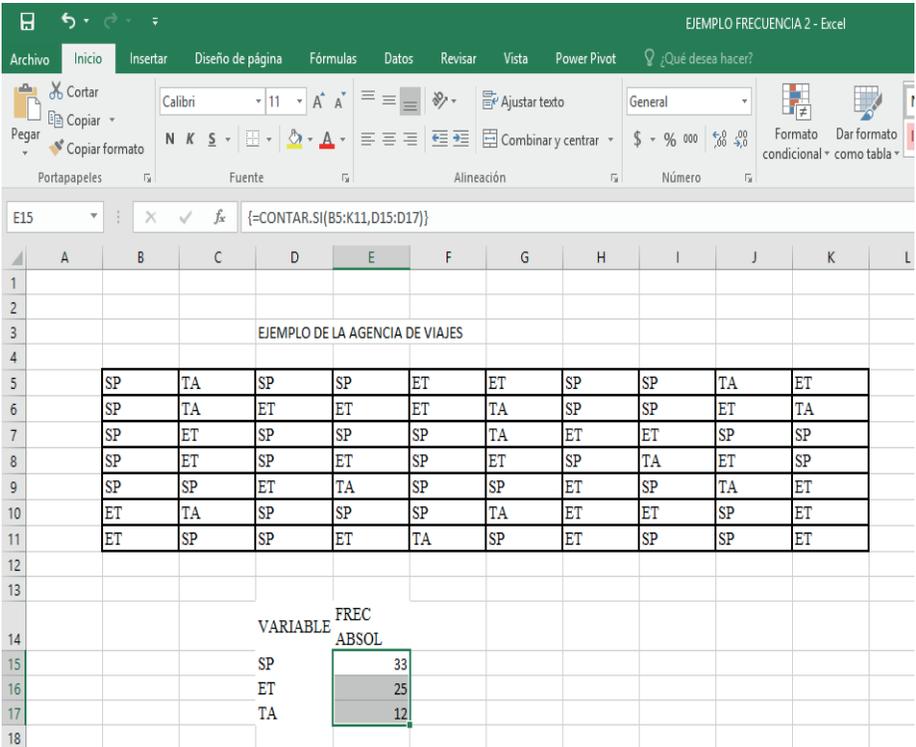


Figura 5. Resultados de la función “CONTAR.SI” con la Frecuencia Absoluta.

Note que el valor coincide con los calculados manualmente. Las demás frecuencias de la Tabla de Distribución de Frecuencias, puede calcularse programándose directamente en la Hoja EXCEL o manualmente como se ha visto anteriormente.

Para el caso que la variable sea cuantitativa continua, la Frecuencia Absoluta se obtiene con el EXCEL utilizando la función estadística “FRECUENCIA”

Para ejemplificar se utilizará los mismos datos del Ejemplo 3 del tiempo de duración del servicio en la caja de un banco (Ivars Escortell, et al., 2005).

Lo primero es colocar los datos del problema y los límites de los intervalos en que se desea agrupar los datos en una hoja EXCEL, tal como se muestra en la Figura 6:

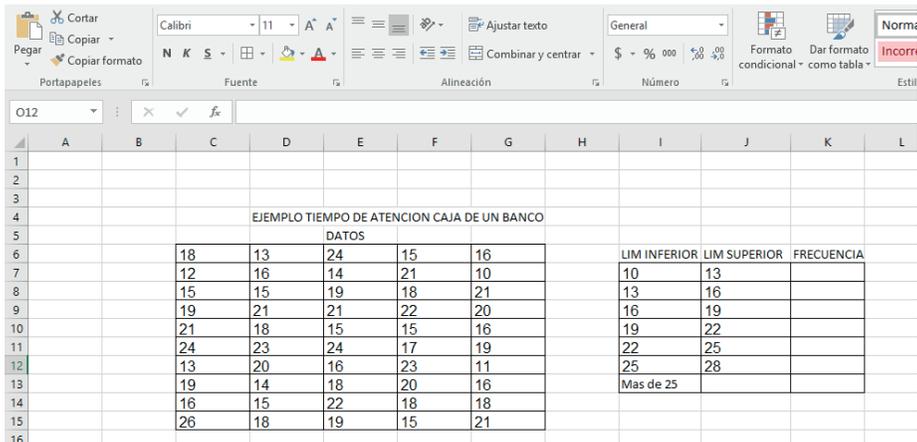


Figura 6. Hoja inicial EXCEL para el cálculo de la Frecuencia Absoluta en una variable cuantitativa. Posteriormente se hace clic en la casilla que está a la derecha del primer límite superior (I3) y se busca la función FRECUENCIA y aparece la ventana de dialogo que se muestra en la Figura 7.

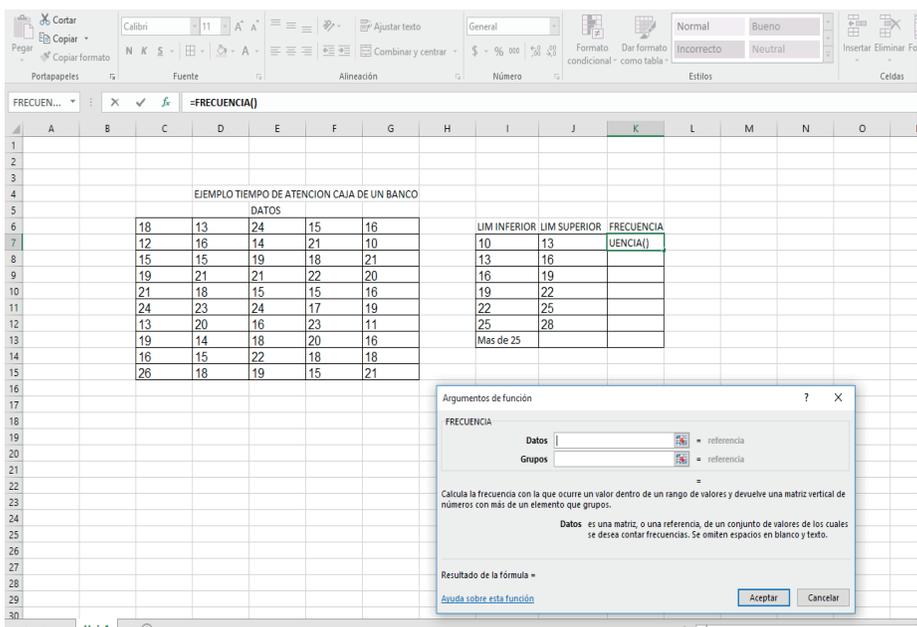


Figura 7. Ventana de la Función FRECUENCIA de EXCEL.

Introduzca los datos en la ventana de la figura 7 y los límites superiores en la ventana “argumento de la función”, tal como se muestra en la Figura 8.

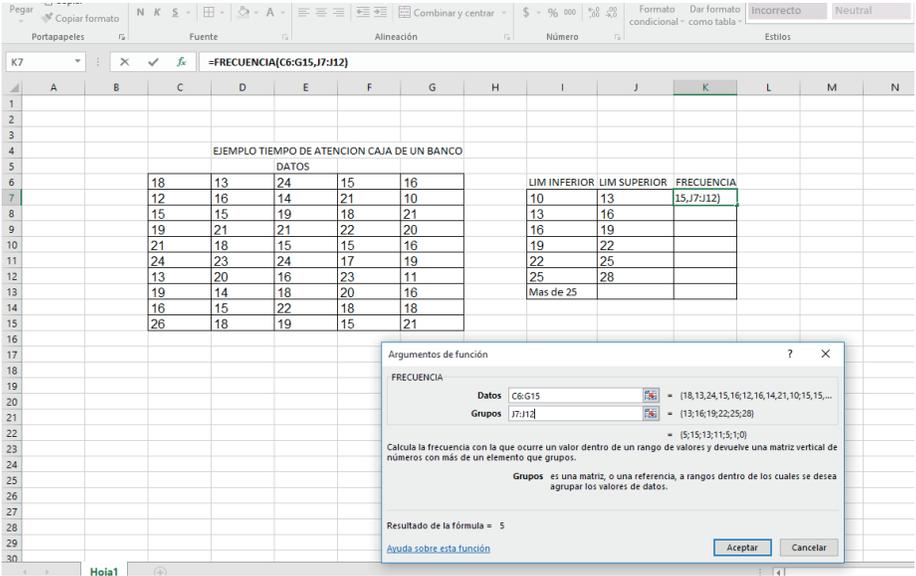


Figura 8. Ventana de la Función FRECUENCIA de EXCEL con datos.

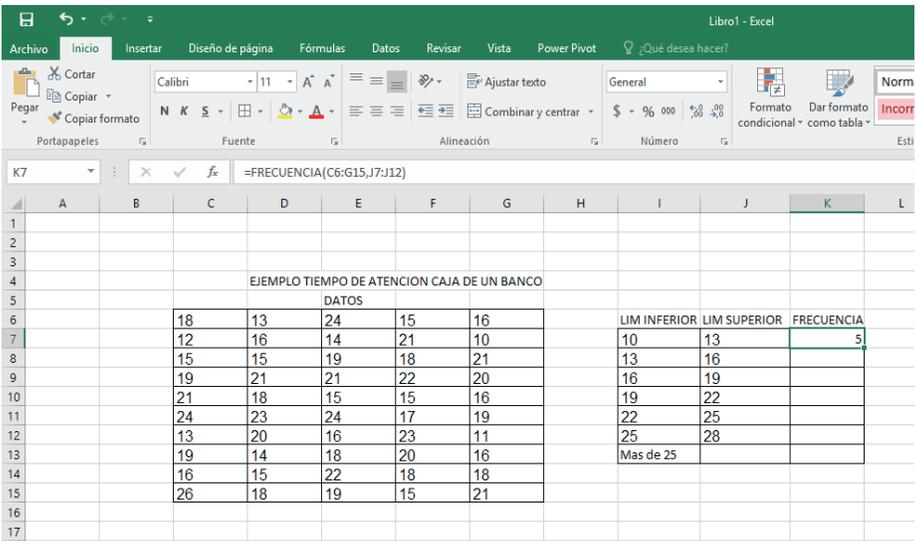


Figura 9. Frecuencia Absoluta para el primer intervalo de la Tabla de Frecuencia.

Finalmente, para encontrar el resto de las frecuencias absolutas, se procede similarmente que para la Función CONTAR.SI. Marque las celdas asociadas a cada intervalo y otra adicional que dará las frecuencias mayores que el último límite superior. Entonces marque F2 y después CONTROL-MAYUSCULA-ENTER y aparecerá el resto de las frecuencias, tal como se muestra en la Figura 10.

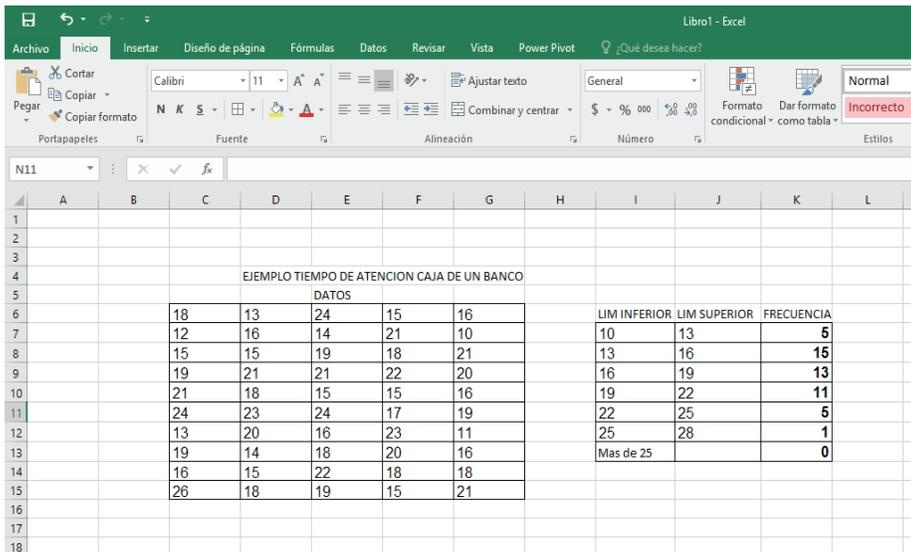


Figura 10. Tabla de Frecuencia Absoluta del Ejemplo 3.

1.5. Problemas resueltos de construcción de tablas de distribución frecuencial

PR 1. En una empresa que produce muebles metálicos para oficinas se evalúan 4 defectos de la calidad del producto: abolladuras (A), pintura (P), ensamblaje (E), cerradura (C). De la producción del pasado mes se tomaron aleatoriamente 60 muebles clasificados como defectuosos para conocer el defecto principal y los resultados se muestran en la Tabla 9:

Tabla 9. Defectos de muebles metálicos para oficinas. Ecuador, 2018.

A	P	E	C	P	A	A	C	P	A
P	A	P	E	E	C	E	P	E	C
P	C	C	P	A	E	E	A	A	E
C	P	A	A	P	C	P	C	A	P

A	A	A	P	A	P	C	E	A	P
E	E	P	A	A	A	A	E	P	A

Fuente: Datos ficticios.

- Encuentre la frecuencia absoluta de forma manual.
- Utilice el EXCEL para calcular la frecuencia absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCIÓN

a) En este problema la variable es cualitativa nominal. El primer paso será realizar un conteo para conocer qué cantidad de muebles de la muestra tomada, presentan los distintos defectos. Esto puede realizarse por cualquier procedimiento de conteo, en este caso utilizaremos un proceso de marcado según la tabla.

ABOLLADURA	//// // // // // // // // // /	21
PINTURA	//// // // // // // // // //	17
ENSAMBLAJE	//// // // // // //	12
CERRADURA	//// // // // //	10
TOTAL		60

b) El paso anterior puede hacerse utilizando el EXCEL como se explicó anteriormente.

Primero se copia los datos en una Hoja de Cálculo EXCEL y se organiza la Tabla de Frecuencia Absoluta en dos columnas, una para el nombre de las variables y la otra para la Frecuencia Absoluta.

Se pone los valores de la variable de estudio y se marca la primera celda de la columna de la Frecuencia Absoluta

Se selecciona la Función CONTAR.SI, ya que la variable es categórica y se rellena la información solicitada en la ventana de esa función. Se muestra en la Figura 11 como queda la Hoja EXCEL después de estos pasos:

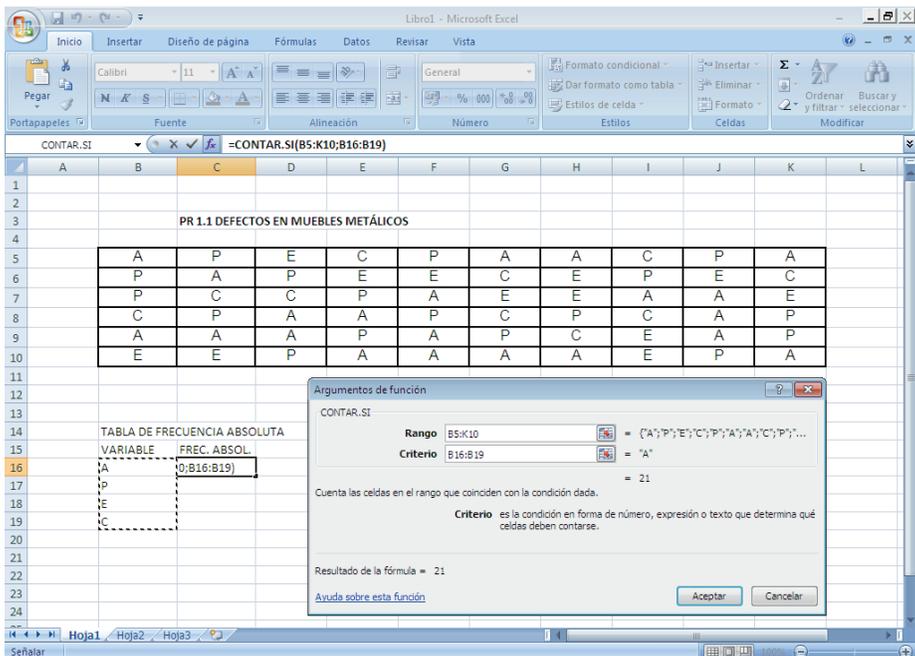


Figura 11. Hoja inicial de EXCEL para el PR1.

Se da aceptar en la función `CONTAR.SI` y aparecerá el valor de la Frecuencia Absoluta para el primer valor de la variable. Para hallar el resto de los valores de la frecuencia, se marcan todas las celdas de la columna de la frecuencia y se da `F2+CONTROL-MAYÚSCULA-ENTER` y aparecerá el resto de los valores de la Frecuencia Absoluta. El resultado se muestra en la Figura 12.

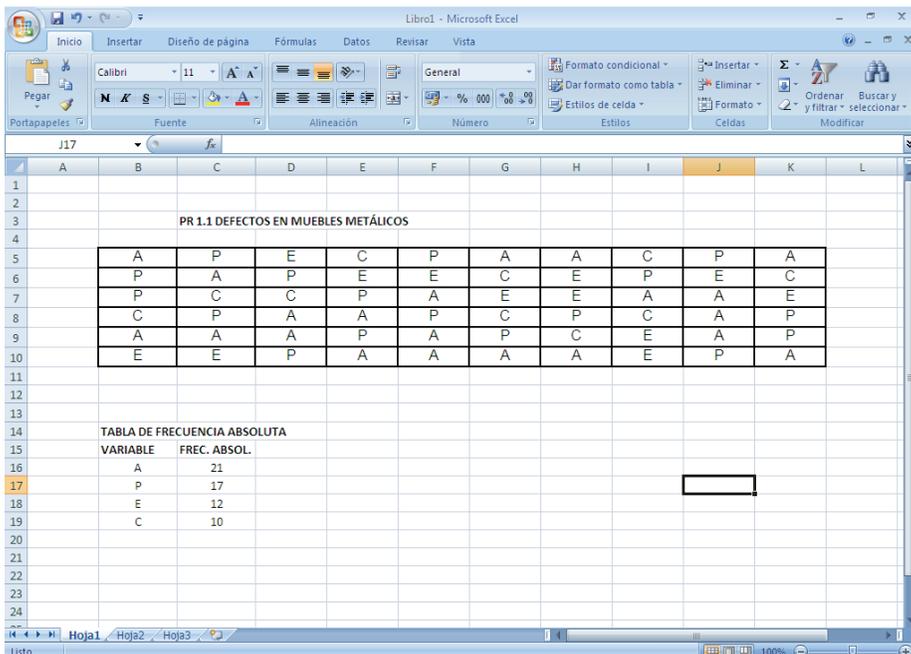


Figura 12. Tabla de Frecuencia Absoluta para PR 1.1

Puede observarse que esta Frecuencia Absoluta coincide con la encontrada de forma manual por el conteo. A partir de aquí, programando en la propia Tabla Excel o de forma manual puede encontrarse el resto de las frecuencias de la Tabla de Distribución de Frecuencias (García & Simón de Blas, 2007).

c) Con esta información puede construirse la Tabla de Distribución de Frecuencia, que tendrá 4 clases, una para cada defecto y los números encontrados en el conteo son la Frecuencia Absoluta para cada clase. A partir de la Frecuencia Absoluta pueden calcularse las distintas columnas de la Tabla 10:

Tabla 10. Distribución de frecuencia “defectos de muebles metálicos” Ecuador, 2018.

DEFECTO	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ABSOLUTA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
ABOLLADURA	21	21	0.35	0.35
PINTURA	17	38	0.28	0.63
ENSAMBLAJE	12	50	0.20	0.83
CERRADURA	10	60	0.17	1.00
TOTAL	60		1.00	

Fuente: Datos ficticios.

La primera columna de la tabla son los números que se encontraron para cada defecto por el conteo de los datos de las 60 observaciones.

La segunda columna, Frecuencia Absoluta Acumulada, se calcula sumando la Frecuencia Absoluta que abarque desde la primera clase, hasta la clase deseada, así:

$$F_{\text{PINTURA}} = 21 + 17 = 38$$

$$F_{\text{ENSAMBLAJE}} = 21 + 17 + 12 = 38 + 12 = 50$$

$$F_{\text{CERRADURA}} = 21 + 17 + 12 + 10 = 50 + 10 = 60$$

Note que el valor de la última frecuencia acumulada coincide con el total de observaciones o datos del problema, en este caso 60.

La tercera columna es el resultado de dividir la Frecuencia Absoluta que aparece en la primera columna, entre el número total de observaciones (Hanke, 1997), esto es:

$$f_{\text{rABOLLADURA}} = 21/60 = 0.35$$

$$f_{\text{rPINTURA}} = 17/60 = 0.28$$

Y así sucesivamente para todas las clases del problema

Por último, los valores de la última columna, la Frecuencia Relativa Acumulada, se calcula sumando la Frecuencia Relativa desde la primera clase hasta la clase deseada, en el problema será:

$$F_{rPINTURA} = 0.35 + 0.28 = 0.63$$

$$F_{rENSAMBLAJE} = 0.35 + 0.28 + 0.20 = 0.63 + 0.20 = 0.83$$

$$F_{rCERRADURA} = 0.35 + 0.28 + 0.20 + 0.17 = 0.83 + 0.17 = 1.00$$

Note que el valor de la Frecuencia Relativa Acumulada en la última clase es igual a 1.00

d) De las distintas columnas de la Tabla puede concluirse que:

- Los defectos más frecuentes son la abolladura y la pintura.
- Si se observa la columna de la Frecuencia Relativa, estos defectos representan respectivamente el 35% y el 28% del total.
- La Frecuencia Relativa Acumulada para estas dos clases es de 0.63, o sea un 63% de los muebles con defectos, presentan alguno de estos dos defectos, lo que implica que deben ser los primeros en analizar para determinar sus causas y tomar las medidas para su solución.

PR 2. Una cafetería que se dedica a la venta de helado quiere estudiar la demanda de los sabores preferidos por los clientes en un fin de semana. Durante cuatro fines de semanas ha tomado una muestra del pedido de 15 clientes y los resultados se muestran en la Tabla 11.

Tabla 11. Sabores preferidos por los clientes en un fin de semana. Ecuador, 2018.

CH	AL	V	AL	CH	AL	FR	CH	AL	FR
V	AL	FR	FR	V	CH	AL	CH	FR	AL
AL	CH	CH	CH	V	CH	CH	FR	AL	CH
CH	CH	CH	CH	CH	AL	AL	V	FR	CH
FR	FR	AL	AL	CH	V	V	CH	AL	FR
AL	V	FR	V	AL	V	FR	CH	CH	AL

Fuente: Datos ficticios.

Leyenda: CH: chocolate, FR: fruta, AL: almendra, V: vainilla

- Encuentre la frecuencia absoluta de forma manual.
- Utilice el EXCEL para calcular la frecuencia absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCIÓN

La variable de este problema es categórica o cualitativa y sus valores se corresponden con los sabores de helados que serán estudiados.

a) Para construir la Tabla de Distribución de Frecuencia primero se realiza un conteo de cada sabor para obtener la frecuencia absoluta de cada uno:

CH	//// // // // // // // /	21
V	//// // //	10
AL	//// // // // // // //	17
FR	//// // // //	12
TOTAL		60

b) Aplicando el EXCEL se puede obtener esta información de forma rápida y segura. Primero se copia los datos en una Hoja de Cálculo EXCEL, se organiza la Tabla de Frecuencia Absoluta en dos columnas, una para el nombre de las variables y la otra para la Frecuencia Absoluta, se pone los valores de la variable de estudio, se marca la primera celda de la columna de la Frecuencia Absoluta, se selecciona la Función CONTAR.SI, ya que la variable es categórica y se rellena la información solicitada en la ventana de esa función. Se muestra en la Figura 13 como queda la Hoja EXCEL después de estos pasos:

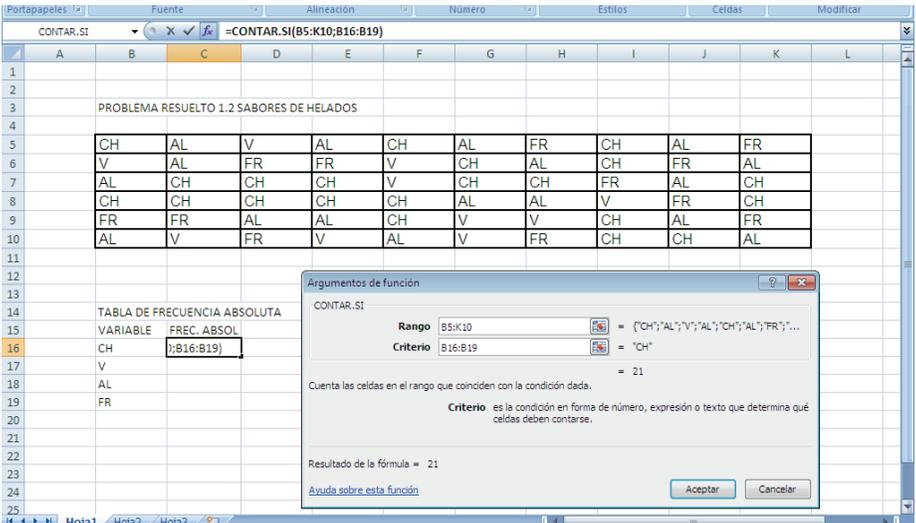


Figura 13. Hoja EXCEL para el cálculo de la Frecuencia Absoluta del PR 2.

Para hallar el resto de los valores de la Frecuencia Absoluta, se marcan todas las celdas de la columna de frecuencia y se selecciona F2 + CONTROL-MAYUSCULA-ENTER y aparecerán estos. La Tabla final se muestra en la Figura 14:

The screenshot shows an Excel spreadsheet with the following content:

- Row 3: PROBLEMA RESUELTO 1.2 SABORES DE HELADOS
- Rows 5-10: A 12x12 grid of ice cream flavor combinations (CH, AL, V, FR).
- Row 14: TABLA DE FRECUENCIA ABSOLUTA
- Row 15: VARIABLE FREC. ABSOL.
- Row 16: CH 21
- Row 17: V 10
- Row 18: AL 17
- Row 19: FR 12
- Row 21: 60

Figura 14. Hoja EXCEL con la Tabla de Frecuencia Absoluta para PR 2.

a) A partir de la Frecuencia Absoluta, se construye la Tabla de Distribución de Frecuencias, la que se muestra en la Tabla 12.

Tabla 12. Distribución de frecuencia de “sabores preferidos” Ecuador, 2018.

SABOR	FREC ABS f_i	FREC ABS ACUMULADA F_i	FREC RELATIVA f_{ri}	FREC REL ACUMULADA F_{ri}
Chocolate	21	21	0.35	0.35
Vainilla	10	31	0.17	0.52
Almendra	17	48	0.28	0.80
Frutas	12	60	0.20	1.00
	60		1.00	

Fuente: Datos ficticios.

d) Análisis de la Tabla 12.

- El sabor más preferido es el chocolate, seguido de almendra.

- Entre los dos representan el 63% de los solicitado por los clientes ($f_{rCH} + f_{rAL} = 0.35 + 0.28 = 0.63$).
- El menos preferido es el sabor de vainilla (17%)

P.R 3. Se quiere estudiar el nivel de desechos de madera en una carpintería y se han tomado 40 observaciones o datos del desecho semanal en kilogramos. Los datos se muestran en la siguiente Tabla 13.

Tabla 13. Frecuencias de nivel de desechos de madera en una carpintería. Ecuador, 2018.

26.3	32.5	23.00	25.2	19.3	21.6	34.0	32.6	30.5	20.8
28.3	35.1	24.5	21.3	27.0	30.1	18.3	21.9	24.4	23.6
22.3	23.0	27.8	32.6	34.4	23.2	29.0	31.4	26.5	28.9
24.6	27.3	30.2	32.5	19.0	23.6	30.0	24.6	27.7	28.0
29.0	21.5	22.4	20.8	27.8	31.1	25.7	22.8	28.6	27.2

Fuente: Datos ficticios.

- Encuentre la frecuencia absoluta de forma manual.
- Utilice el EXCEL para calcular la frecuencia absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCIÓN

- Los datos mostrados en la tabla se corresponden con una variable cuantitativa continua. El primer paso es determinar el rango en que se mueven los datos.

$$\text{RANGO} = \text{MAYOR VALOR} - \text{MENOR VALOR} = 35.1 - 18.3 = 16.8 \text{ kg}$$

- Para construir los intervalos debe determinarse el límite inferior y superior de cada intervalo y debe cumplirse que los intervalos determinados contengan todas las observaciones. La amplitud del intervalo es el rango dividido entre el número de intervalos que se desea, suponga para este problema que sean 5 intervalos (Valdivieso, 2006).

$$\text{AMPLITUD INTERVALO} = \text{RANGO} / \text{NUMERO DE INTERVALOS} = 16.8 / 5 = 3.4 \text{ kg}$$

Para construir los intervalos, se iniciará por el valor mínimo de los datos y se sumará la amplitud calculada. De esta forma los intervalos quedarán formados como:

[18.3 - 21.7] “nótese, que al límite inferior se le suma la amplitud del intervalo, es decir, 3.4 y se obtiene 21.7, esta operación se realiza en cada intervalo”.

(21.7- 25.1]

(25.1 - 28.5]

(28.5 - 31.9]

(31.9 - 35.3]

Contando los datos que caen en cada intervalo se tiene:

[18.3 - 21.7]	//// //	8
(21.7 - 25.1]	//// // // //	13
(25.1 - 28.5]	//// // // //	12
(28.5 - 31.9]	//// // //	10
(31.9 - 35.3]	//// //	7

b) Esto puede resolverse utilizando el EXCEL. Para ello se coloca en una hoja EXCEL los datos, los intervalos y el límite superior de cada intervalo, tal como se muestra en la siguiente Figura 15:

The screenshot shows an Excel spreadsheet with the following content:

PROBLEMA RESUELTO 1.3 DESECHOS DE MADERA

26.3	32.5	23	25.2	19.3	21.6	34	32.6	30.5	20.8
28.3	35.1	24.5	21.3	27	30.1	18.3	21.9	24.4	23.6
22.3	23	27.8	32.6	34.4	23.2	29	31.4	26.5	28.9
24.6	27.3	30.2	32.5	19	23.6	30	24.6	27.7	28
29	21.5	22.4	20.8	27.8	31.1	25.7	22.8	28.6	27.2

DISTRIBUCION DE FRECUENCIA ABSOLUTA

INTERVALO	LIMITE SUPERIOR	FRECUENCIA ABSOLUTA
18.3 - 21.7	21.7	
21.7 - 25.1	25.1	
25.1 - 28.5	28.5	
28.5 - 31.5	31.5	
31.5 - 35.3	35.3	

Figura 15. Hoja EXCEL inicial para los desechos de madera.

Se marca la primera fila de la columna de la Frecuencia Absoluta y se selecciona la Función FRECUENCIA y aparece una ventana donde se solicita datos para calcular la frecuencia. Donde dice Datos se selecciona la tabla de datos del problema y en la casilla que dice Grupos se colocan los límites superiores de cada intervalo. Esto se muestra en la Figura 16:

The screenshot shows an Excel spreadsheet with the following data:

PROBLEMA RESUELTO 1.3 DESECHOS DE MADERA										
26,3	32,5	23	25,2	19,3	21,6	34	32,6	30,5	20,8	
28,3	35,1	24,5	21,3	27	30,1	18,3	21,9	24,4	23,6	
22,3	23	27,8	32,6	34,4	23,2	29	31,4	26,5	28,9	
24,6	27,3	30,2	32,5	19	23,6	30	24,6	27,7	28	
29	21,5	22,4	20,8	27,8	31,1	25,7	22,8	28,6	27,2	

INTERVALO	LIMITE SUPERIOR	FRECUENCIA ABSOLUTA
18.3 - 21.7	21,7	=FREC(C14:C18)
21.7 - 25.1	25,1	
25.1 - 28.5	28,5	
28.5 - 31.9	31,9	
31.9 - 35.3	35,3	

The dialog box for the FRECUENCIA function shows the following configuration:

- Datos:** B5:K9
- Grupos:** C14:C18
- Resultado de la fórmula:** = 8

Figura 16. Hoja EXCEL con los datos para el cálculo de la Frecuencia Absoluta de los desechos de madera.

Se da aceptar y aparece la Frecuencia Absoluta para el primer intervalo. Para hallar el resto de las frecuencias, se procede similarmente a lo hecho para la Función CONTAR.SI, esto es, se marca toda la columna de la Frecuencia Absoluta y se selecciona F2+CONTROL-MAYUSCULA-ENTER y aparece el resto de las frecuencias. Esto se muestra en la siguiente Figura 17.

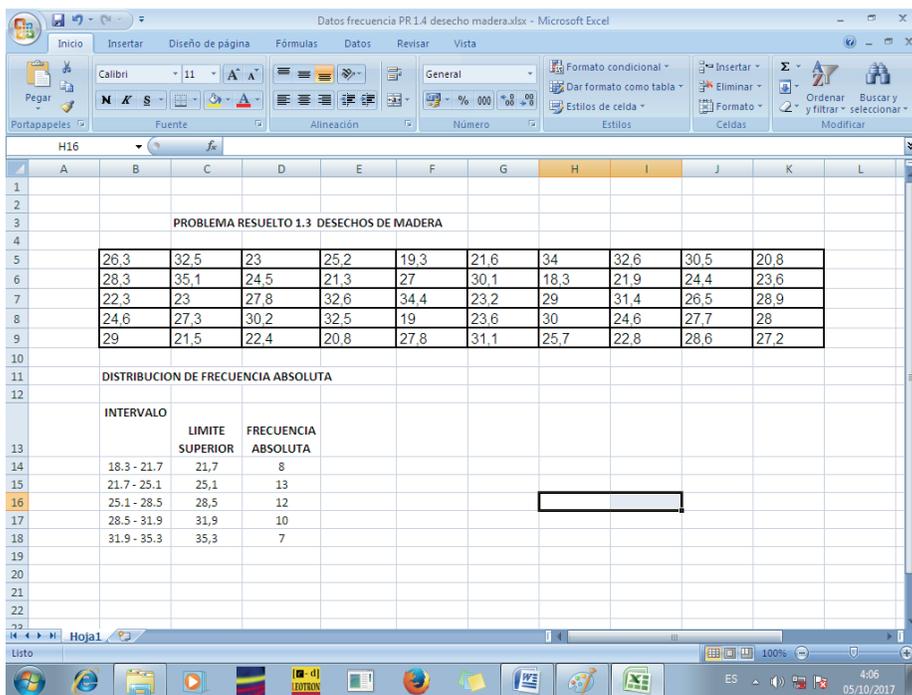


Figura 17. Hoja EXCEL con la Frecuencia Absoluta de los desechos de madera.

b) De forma manual o sobre la misma hoja EXCEL, puede construirse la Tabla de Distribución de Frecuencias que se muestra en la Tabla 14.

Tabla 14. Distribución de frecuencias de los desechos de madera. Ecuador, 2018.

INTERVALOS PARA EL DESECHO EN KG/SEMANA	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
(18.3 - 21.7]	8	8	0.16	0.16
(21.7- 25.1]	13	21	0.26	0.42
(25.1 - 28.5]	12	33	0.24	0.66
(28.5 - 31.9]	10	43	0.20	0.86
(31.9 - 35.3]	7	50	0.14	1.00
TOTAL	50		1.00	

Fuente: Datos ficticios.

d) Análisis de la Tabla de Distribución de Frecuencias

- Puede concluirse que la cantidad de desechos de madera semanal más frecuente, está en los intervalos de 21.7 a 25.1 kg y de 25.1 a 28.5 kg con un 26% y 24% respectivamente, o sea un 50% de las semanas el desecho de madera de la carpintería estará entre 21.7 y 28.5 kg.
- Solo un 16% de las semanas el desecho de madera estará por debajo de 21.7 kg y un 14% por encima de 31.9 kg
- Un 86% de las semanas el desecho de madera de la carpintería será de 31.9 kg o menos.
- Un 68% de las semanas el desecho de madera será menor o igual a 28.5 kg.

PR 4. Se está realizando un estudio de la circulación vial por una carretera y se han tomado 50 horas de muestra para medir la cantidad de autos que circula por hora. Los resultados se muestran en la Tabla 15.

Tabla 15. Frecuencia de la circulación de autos por horas en una vía. Ecuador, 2018.

22	25	21	23	20	21	23	23	20	24
20	24	22	21	21	24	24	25	23	23
21	22	25	20	24	25	20	21	24	24
20	23	25	23	23	23	21	20	23	21
23	20	24	24	22	22	24	24	25	22

Fuente: Datos ficticios

- a) Encuentre la frecuencia absoluta de forma manual.
- b) Utilice el EXCEL para calcular la frecuencia absoluta de este problema.
- c) Construya una Tabla de Distribución de Frecuencias.
- d) Analice los resultados.

SOLUCION

- a) Nótese que esta variable es cuantitativa discreta, por lo que, de acuerdo al interés del investigador, puede organizarse por categorías o clases (cada valor de la variable) o por intervalos. En este ejemplo se hará por las dos formas (Coscolluela, Turbany & Fornieles, 2002; Christensen, 2012)

Primero se realiza un conteo en este caso organizado por clases, o sea por cada valor que tiene la variable. Los resultados son:

20	//// //	8
21	//// //	8
22	//// /	6
23	//// //// /	11
24	//// //// /	11
25	//// /	6
TOTAL		50

b) Utilizando el EXCEL, se emplearía la Función CONTAR.SI y repitiendo los pasos vistos en los dos primeros problemas resueltos, quedaría la Distribución de la Frecuencia Absoluta que se muestra en la Figura 1.18

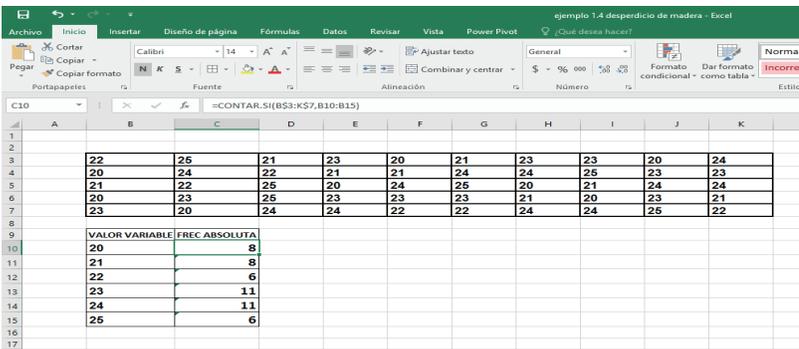


Figura 18. Hoja EXCEL con la Frecuencia Absoluta para PR 3.

Si se hubiera organizado por intervalos, se seguiría el mismo procedimiento que en el problema resuelto PR 3.

Para este caso debido a los pocos valores de las variables, de 20 a 26, se pueden construir 3 intervalos: de 20 a 22, de 22 a 24 y de 24 a 26, lo cual limita los análisis a realizar.

La Frecuencia Absoluta es calculada a partir de los resultados del inciso anterior. Asumiendo que valores coincidentes con los límites se cuenta en el límite superior, esta será:

INTERVALO	FRECUENCIA ABSOLUTA
20 - 22	$8+8+6 = 22$

22 - 24	11+11 = 22
24 - 26	6

Tabla 17. Distribución de Frecuencias autos por horas. Ecuador, 2018.

CANTIDAD VEHICULOS / HORA	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
20 - 22	22	16	0.32	0.32
22 - 24	22	33	0.34	0.66
24 - 26	6	50	0.34	1.00
TOTAL	50		1.00	

Fuente: Datos ficticios.

A partir de la Frecuencia Absoluta calculada para cada clase de la variable, se construye la Tabla de Distribución de Frecuencias, que para este problema se muestra en la Tabla 18.

Tabla 18. Distribución de Frecuencias por valores de la variable de autos por horas. Ecuador, 2018.

CANTIDAD VEHICULOS /HORA	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
20	8	8	0.16	0.16
21	8	16	0.16	0.32
22	6	22	0.12	0.44
23	11	33	0.22	0.66
24	11	44	0.22	0.88
25	6	50	0.12	1.00
TOTAL	50		1.00	

Fuente: Datos ficticios

d) A partir de la Tabla de Distribución de Frecuencias en clases se puede observar que:

- Las mayores frecuencias relativas y por tanto los mayores porcentos, se corresponden con los valores de 23 y 24 vehículos por hora con un 22% para cada una de ellas.
- Para el valor de 25 autos por hora la frecuencia relativa es de un 12%, igual que la de 22 vehículos por hora que son las más bajas.
- Para 24 o menos autos hay una frecuencia relativa acumulada de un 0.88, o sea que un 88% de las horas observadas el tráfico vehicular fue de 24 autos o menos.

PR 5. En la línea de envasado de una planta de producción de café molido, los paquetes de 250 gramos, deben tener un peso entre 249.90 y 250.10 gramos. Para comprobar este índice de calidad, se ha tomado una muestra de 80 paquetes de café al final de la línea y los resultados se muestran en la Tabla 19.

Tabla 19. Peso de un paquete de café molido. Ecuador, 2018

250.00	249.95	249.98	250.03	250.08	249.98	249.91	250.10	249.95	250.09
249.96	250.02	250.00	249.97	250.08	250.12	249.95	250.09	250.00	249.99
250.00	249.86	250.10	250.04	249.98	249.96	250.00	250.00	250.03	249.97
250.12	250.00	250.02	250.00	249.99	249.98	249.96	250.00	250.09	250.00
250.03	249.93	249.90	250.09	250.02	249.98	249.96	250.14	250.04	249.93
249.90	250.12	250.00	249.99	250.02	250.00	250.00	249.87	250.01	250.00
250.00	250.09	250.04	249.95	250.00	250.01	249.98	250.00	250.01	249.96
250.00	249.99	250.00	250.08	249.90	249.93	250.06	250.00	250.01	249.91

Fuente: Datos ficticios.

- Utilice el EXCEL para calcular la frecuencia absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCION

- Los datos suministrados en la tabla se corresponden con una variable aleatoria cuantitativa continua. El primer paso para poder determinar los límites de los intervalos será determinar el rango en que se mueven esos datos.

$$\text{RANGO} = \text{MAYOR VALOR} - \text{MENOR VALOR} = 250.14 - 249.86 = 0.28$$

Para agrupar los datos en 5 intervalos, la amplitud del intervalo será aproximadamente de 0.06. Si se inicia en el valor de 249.86, entonces los intervalos son:

LIM INF LIM SUP

249.86	249.92
249.92	249.98
249.98	250.04
250.04	250.10
250.10	250.16

Aplicando directamente el EXCEL para el cálculo de la Frecuencia Absoluta, se pondrá primeramente los datos en una hoja EXCEL y los límites para los intervalos determinados anteriormente, tal como se muestra en la Figura 19:

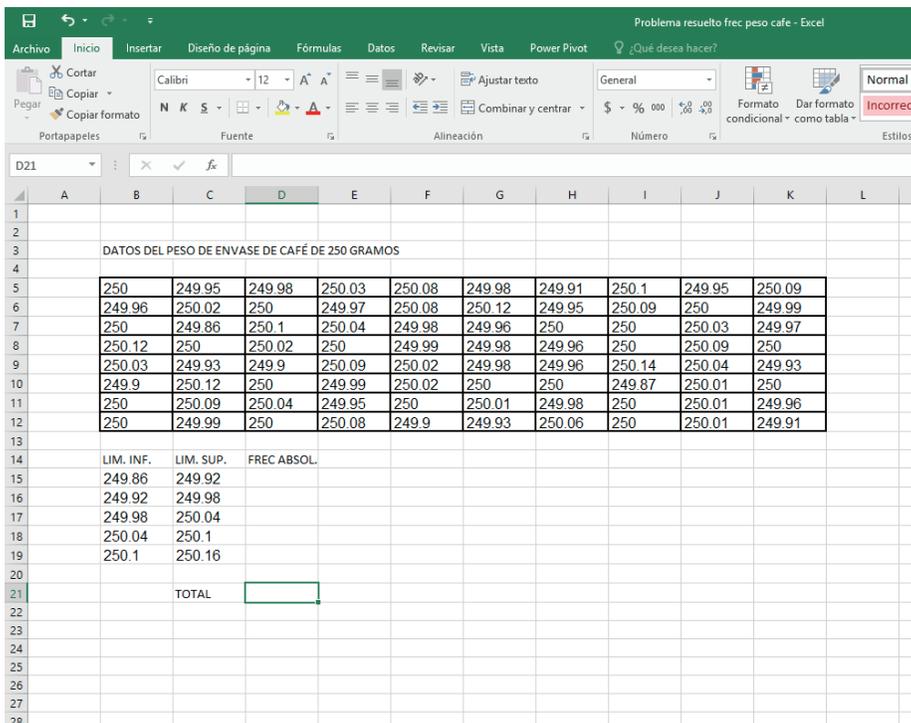


Figura 19. Hoja EXCEL inicial para PR 5.

Como la variable en estudio es cuantitativa continua se hará uso de la función FRECUENCIA.

En datos se cargará la matriz de datos del problema y en grupos los límites superiores de los intervalos determinados y aparece la Frecuencia Absoluta de los 5 intervalos y otro referido a la frecuencia para valores mayores que el último límite superior, en este caso cero, lo que se muestra en la Figura 20:

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2																
3				DATOS DEL PESO DE ENVASE DE CAFÉ DE 250 GRAMOS												
4																
5		250	249.95	249.98	250.03	250.08	249.98	249.91	250.1	249.95	250.09					
6		249.96	250.02	250	249.97	250.08	250.12	249.95	250.09	250	249.99					
7		250	249.86	250.1	250.04	249.98	249.96	250	250	250.03	249.97					
8		250.12	250	250.02	250	249.99	249.98	249.96	250	250.09	250					
9		250.03	249.93	249.9	250.09	250.02	249.98	249.96	250.14	250.04	249.93					
10		249.9	250.12	250	249.99	250.02	250	250	249.87	250.01	250					
11		250	250.09	250.04	249.95	250	250.01	249.98	250	250.01	249.96					
12		250	249.99	250	250.08	249.9	249.93	250.06	250	250.01	249.91					
13																
14		LIM. INF.	LIM. SUP.	FREC ABSOL.												
15		249.86	249.92	15:C19												
16		249.92	249.98													
17		249.98	250.04													
18		250.04	250.1													
19		250.1	250.16													
20																
21				TOTAL												
22																
23																
24																
25																
26																
27																
28																
29																

The dialog box for the FRECUENCIA function is open, showing the following details:

- Argumentos de función:** FRECUENCIA
- Datos:** B5:K12 = {250,249.95,249.98,250.03,250.08,249.98,249.91,250.1,249.95,250.09,250,249.99}
- Grupos:** C15:C19 = {249.92,249.98,250.04,250.1,250.16}
- Resultado de la fórmula:** = {7,20,28,11,4,0}

Figura 20. Hoja EXCEL con los datos para calcular la Frecuencia Absoluta del PR 5.

Hay que recordar que en la hoja EXCEL, en la celda donde se marcó la función aparecerá solo la frecuencia para el primer intervalo y para el resto de los intervalos, se debe marcar el resto de las celdas asociadas a cada intervalo, y dar F2, CONTROL-MAYUSCULA-ENTER y aparecerán las otras frecuencias absolutas, lo que se muestra en la Figura 21.

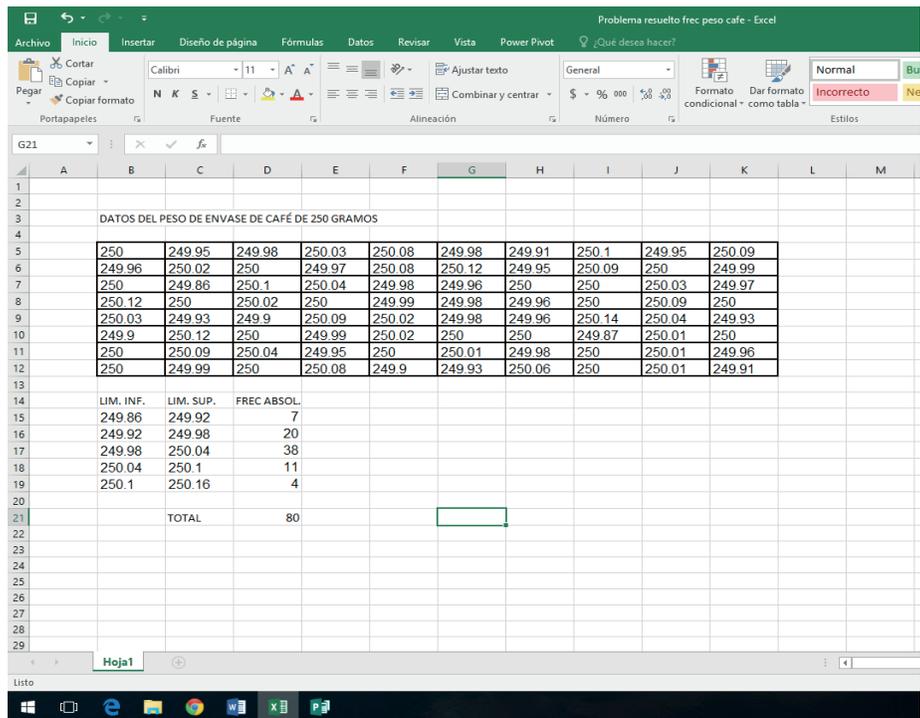


Figura 21. Hoja EXCEL con la Frecuencia Absoluta del PR 5.

b) Con esta información puede construirse el resto de la Tabla de Distribución de Frecuencias, tal como se muestra en la Tabla 20.

Tabla 20. Distribución de frecuencia de pesos de paquete de café molido. Ecuador, 2018.

INTERVALOS PARA EL PESO DE LOS ENVASES DE CAFÉ (GRAMOS)	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
[249.86-249.92]	7	7	0.0875	0.0875
(249.92-249.98]	20	27	0.2500	0.3375
(249.98-250.04]	38	65	0.4750	0.8126
(250.04-250.10]	11	76	0.1375	0.9500
(250.10-250.16]	4	80	0.0500	1.0000
TOTAL	80		1.00	

Fuente: Datos ficticios.

c) Análisis de la tabla de frecuencias

- Un 47,50% de los paquetes de café tienen un peso entre 249.98 gramos y 250.04 gramos, bastante cercano al peso de 250 gramos.
- Un 86.25% de los paquetes de café está entre 249.92 y 250.10 gramos los que cumplen con las especificaciones del peso que tiene la empresa.
- Por encima del límite superior de especificación del peso de un paquete (250.10 gramos) solo hay un 5% de los paquetes producidos.
- En el intervalo de 249.86 a 290.2 gramos, de las 7 observaciones que aparecen marcadas en la tabla de datos, solo 2 no cumplen con la especificación de peso mínimo, o sea solo hay un 7.5% $(4 + 2)/80$ que no cumplen con las especificaciones de peso para un paquete de café de 250 gramos que tiene fijada la empresa.

PR 6. Se ha realizado una prueba diagnóstica sobre 100 puntos, para medir el nivel de entrada en matemática de los estudiantes que ingresan a una universidad. Para ello se han seleccionado al azar 60 estudiantes y las notas obtenidas se muestran en la siguiente Tabla 21.

Tabla 21. Datos sobre prueba diagnóstica.

65	41	54	63	86	67	93	71	63	57	77	59
69	66	74	56	75	61	95	84	73	73	81	58
66	57	66	74	58	73	99	39	70	57	68	59
75	41	57	48	82	67	98	91	60	73	59	65
77	56	44	75	90	95	48	53	91	72	63	81

Fuente: Datos ficticios.

Con esos datos:

- Utilice el EXCEL para calcular la Frecuencia Absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCION

a) Para calcular la Frecuencia Absoluta utilizando el EXCEL, es necesario definir los intervalos en que se agruparan los datos.

Mínimo valor = 39

Máximo valor = 99

Rango = 99 – 39 = 60

Si se selecciona 6 intervalos para agrupar los datos, entonces:

Amplitud del Intervalo = Rango / No. de intervalos = 60 / 6 = 10 puntos

Por tanto, los intervalos serán:

39 – 49; 49 – 59; 59 - 69; 69 – 79; 79 – 89; 89 – 99

En una hoja de cálculo EXCEL se copia la tabla de datos y se ponen los valores del límite superior de los intervalos, se selecciona la Función FRECUENCIA, pues la variable es numérica o cuantitativa y se introduce la información en las ventanas de la función. Finalmente, esta Hoja EXCEL se muestra en la Figura 22.

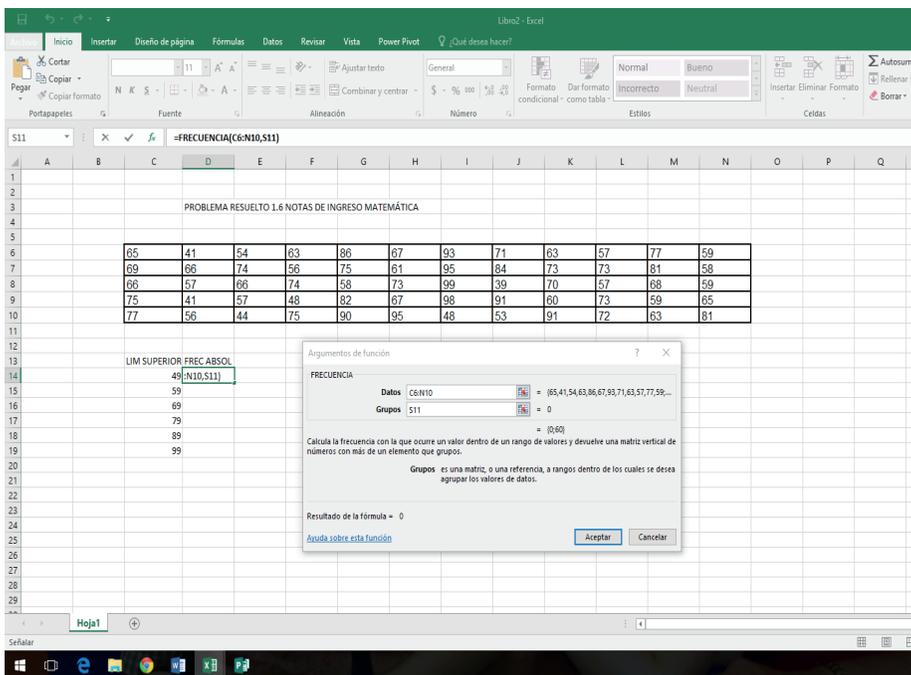


Figura 22. Hoja EXCEL para el cálculo de la Frecuencia Absoluta del PR 6.

Se marca "Aceptar" y aparece la frecuencia en el primer intervalo, de 39 a 49 puntos. Se marca el resto de las celdas de la columna de FRECUENCIA

ABSOLUTA y se selecciona F2, CONTROL-MAYUSCULA-ENTER y aparecerá el resto de las frecuencias absolutas para cada intervalo. Finalmente, sobre la misma hoja EXCEL puede construirse la Tabla de Distribución de Frecuencia, lo que se muestra en la Figura 23:

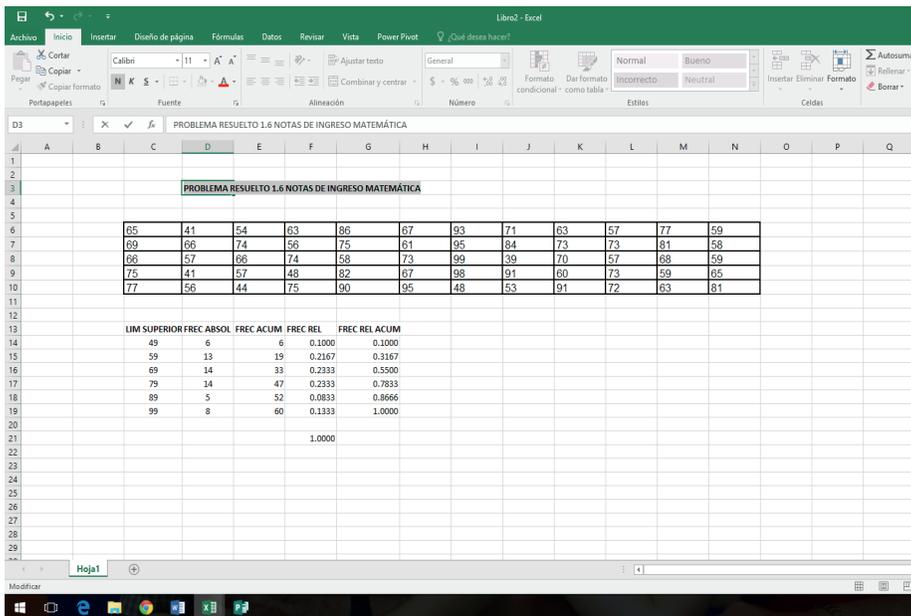


Figura 23. Hoja EXCEL con la Tabla de Distribución de Frecuencias del PR 6.

c) A partir de la Tabla de Distribución de Frecuencias, se pueden sacar las siguientes conclusiones:

- En la columna de la Frecuencia Relativa Acumulada se observa que un 55 % de los alumnos examinados está por debajo de los 69 puntos. Si se considera 70 puntos como aprobado, esto implica que más de la mitad de los estudiantes examinados están desaprobados. Aproximadamente un 32% obtiene notas inferiores a 60 puntos.
- En la columna de la Frecuencia Relativa Absoluta, se puede observar que solo un 13% de los estudiantes examinados obtienen notas de más de 90 puntos.

Puede concluirse que el resultado del examen diagnóstico de matemática no resulta positivo.

PR 7. Una Cooperativa de transporte de pasajeros por ómnibus interprovincial quiere conocer la evaluación que tienen los clientes de la calidad de su servicio. Para ello ha realizado una encuesta a 60 clientes escogidos al azar y el resultado obtenido se muestra en la Tabla 22.

Tabla 22. Resultados de la encuesta.

MB	B	R	B	B	R	MB	M	R	P	B	R
E	R	B	B	P	M	M	R	MB	MB	B	R
B	MB	R	P	M	R	M	MB	E	MB	R	B
R	R	E	MB	R	P	M	M	MB	E	B	B
R	MB	E	B	B	R	M	P	R	B	B	MB

Fuente: Datos ficticios.

Donde: Excelente (E); Muy Bueno (MB); Bueno (B); Regular (R); Malo (M); Pésimo (P).

Con los datos anteriores:

- Encuentre la frecuencia absoluta.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCION

En este problema la variable es de tipo cualitativa. Para determinar la Frecuencia Absoluta se aplicará directamente el EXCEL. Primeramente, la Tabla de datos se copiará en una Hoja EXCEL y se aplica la función CONTAR. SI. Esta hoja de cálculo inicial se muestra en la Figura 24.

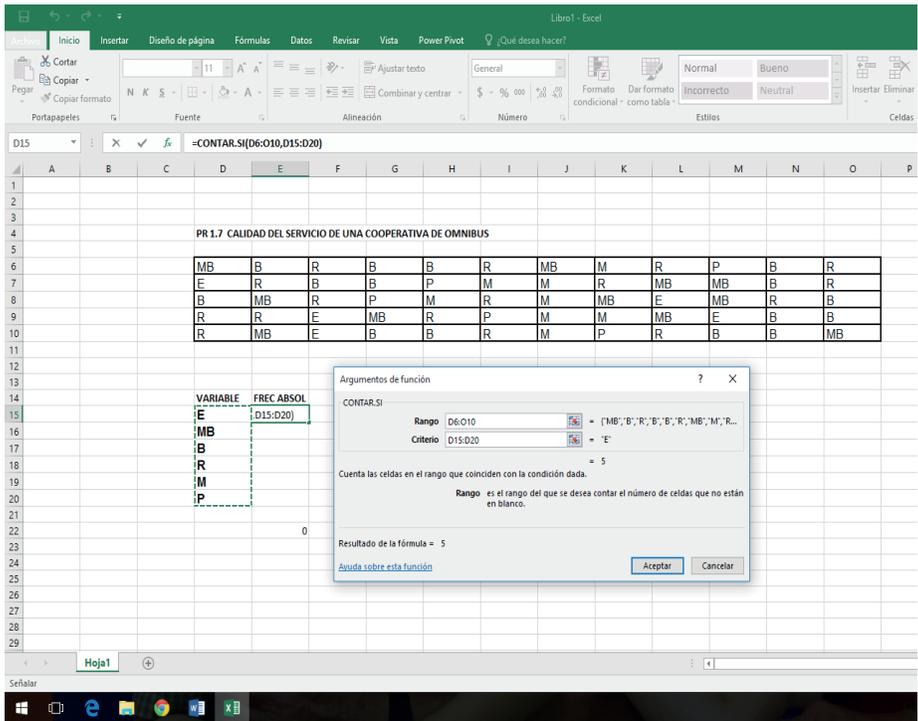


Figura 24. Hoja EXCEL para el cálculo de la Frecuencia Absoluta del PR 7.

Posteriormente se da Aceptar y aparece la frecuencia en la casilla correspondiente a la categoría E, se marcan todas las restantes casillas de esa columna y se da F2, CONTROL-MAYUSCULA-ENTER Y aparece al resto de las Frecuencias Absolutas. En la misma hoja EXCEL, se ha construido la Tabla de Distribución de Frecuencias, tal como se muestra en la Figura 25:

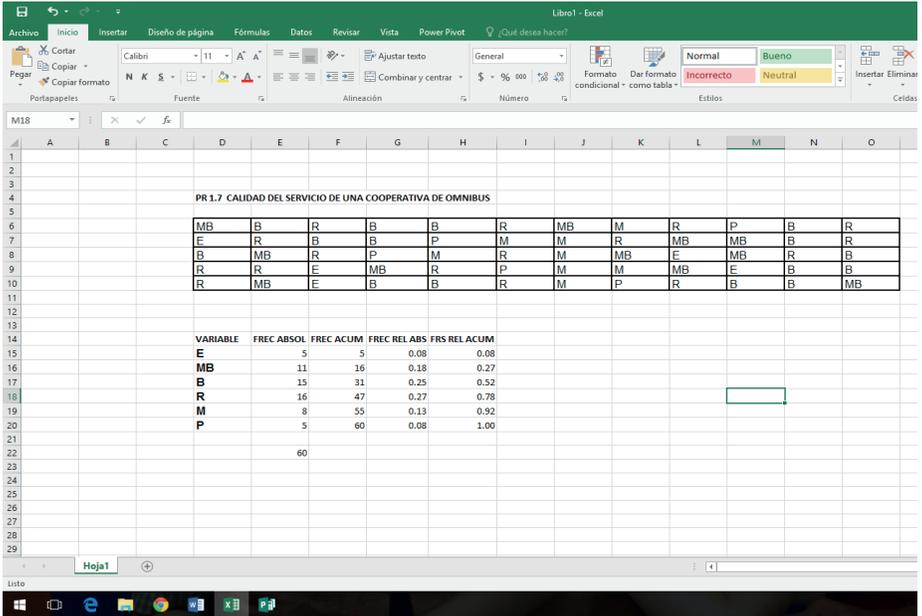


Figura 25. Hoja EXCEL con la Tabla de Distribución de Frecuencias.

c) De la Tabla de Distribución de Frecuencias puede concluirse:

- En la columna de la Frecuencia Relativa Acumulada se observa que solo un 52% evalúa el servicio entre Excelente y Bien y por ende un 48% lo considera Regular, Mal o Pésimo. Solo un 8% evalúa el servicio de Excelente.
- Un 27%, el más alto, lo evalúa de Regular.
- Entre Mal y Pésimo lo evalúan un 21% de los clientes.

Con esta información puede concluirse que es necesario profundizar en las causas que afectan este servicio para mejorar su calidad.

PR 8. Una empresa camaronera tiene contratada la entrega de camarón talla mediana (41 a 50 camarones por libra). Se ha realizado un muestreo en una piscina tomando 60 libras y contando el número de camarones. Estos datos se muestran en la Tabla 23.

Tabla 23. Datos talla de camarón. Ecuador, 2018.

43	48	46	52	41	40	38	40	46	43	44	50
51	43	44	49	39	40	48	41	44	51	39	45
52	50	46	41	49	50	43	47	45	48	41	39
46	49	52	43	50	47	46	44	47	52	39	45
39	48	41	40	46	44	50	43	46	48	43	40

Fuente: Datos ficticios.

Con esos datos:

- Utilice el EXCEL para calcular la Frecuencia Absoluta de este problema.
- Construya una Tabla de Distribución de Frecuencias.
- Analice los resultados.

SOLUCION

Los datos de este problema son de una variable numérica discreta, por lo que podemos resolverlo determinado la frecuencia para los distintos valores o agrupándolo por intervalos. Se resolverá para ambos casos.

Si se asume cada valor de la variable por separado se utiliza la función CONTAR.SI de Excel. Para ello se copia en una Hoja de Cálculo los datos del problema, se selecciona la función CONTAR.SI y se llenan los espacios de la ventana correspondiente a esa función, tal como se muestra en la Figura 26.

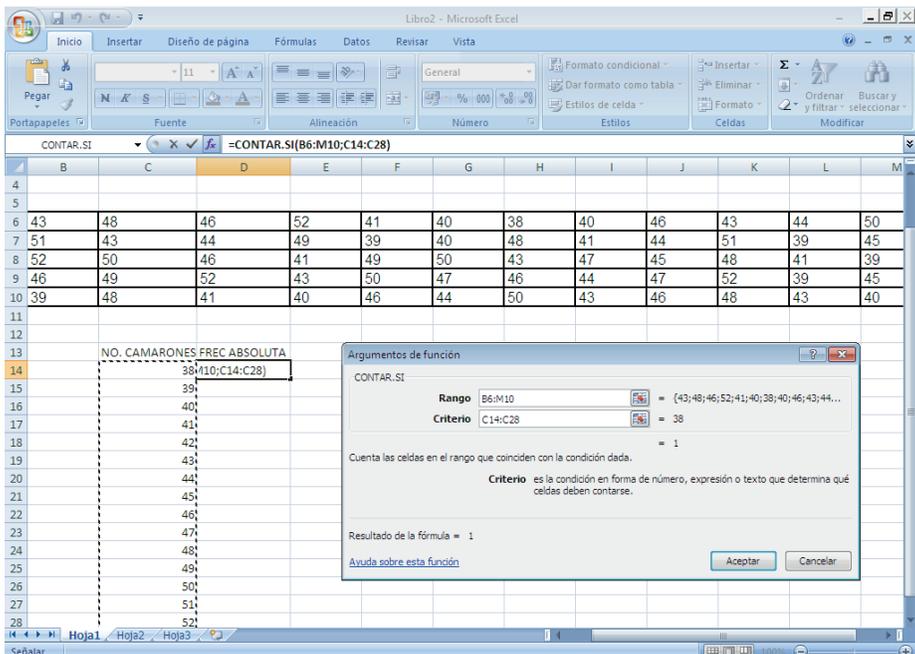


Figura 26. Hoja EXCEL para la aplicación de la Función CONTAR.SI al PR 8.

Posteriormente se da Enter y aparecerá el valor de la Frecuencia Absoluta en la fila correspondiente al valor de 38. Marcando el resto de las celdas de la columna de Frecuencia y dando F2, CONTROL-MAYUSCULA-ENTER aparecerá el resto de las Frecuencias Absolutas y a partir de ellas se puede construir la Tabla de Frecuencias para este problema, tal como se muestra en la Tabla 24.

Tabla 24. Tabla de frecuencias para el PR 8 talla de camarón.

NO. CAMARONES	FREC ABSOLUTA	FREC ACUM	FREC REL ABS	FREC REL ACUM
38	1	1	0,0167	0,0167
39	5	6	0,0833	0,1000
40	5	11	0,0833	0,1833
41	5	16	0,0833	0,2667
42	0	16	0,0000	0,2667
43	7	23	0,1167	0,3833
44	5	28	0,0833	0,4667

45	3	31	0,0500	0,5167
46	7	38	0,1167	0,6333
47	3	41	0,0500	0,6833
48	5	46	0,0833	0,7667
49	3	49	0,0500	0,8167
50	5	54	0,0833	0,9000
51	2	56	0,0333	0,9333
52	4	60	0,0667	1,0000

Fuente: Datos ficticios.

Para el caso que se desee agrupar los datos por intervalo, primero será necesario definir los intervalos.

Mayor valor = 52

Menos valor = 38

Rango = $52 - 38 = 14$

Se agruparán los datos en 7 intervalos y por tanto la amplitud del intervalo es:

$$14 / 7 = 2$$

Y los intervalos son los siguientes:

38 – 40; 40 – 42; 42 – 44; 44 – 46; 46 – 48; 48 – 50 y de 50 – 52

Para hallar la Frecuencia Absoluta en cada intervalo se utiliza la función FRECUENCIA de EXCEL. Para ello se construye una tabla en la Hoja de Cálculo donde se ponen los intervalos, el límite superior de cada intervalo y la Frecuencia Absoluta. Situado en la primera celda de la Frecuencia se escoge la función FRECUENCIA y se llenan los distintos espacios de la ventana de esa función, tal como se muestra en la Figura 27:

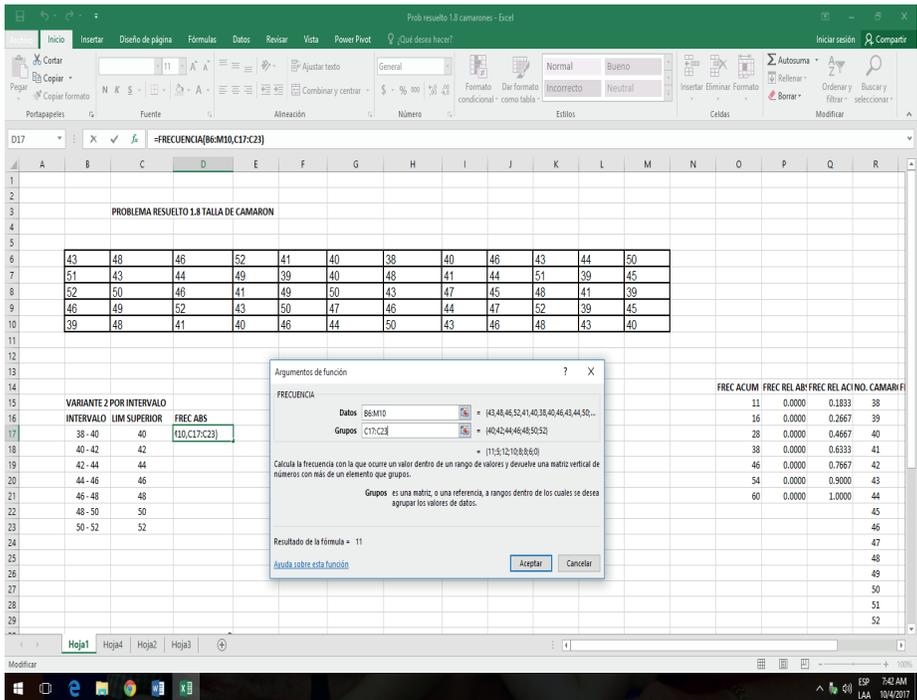


Figura 27. Hoja EXCEL para el uso de la Función FRECUENCIA del PR 8.

Se da en Aceptar y aparecerá la Frecuencia Absoluta para el primer intervalo y como ya se ha explicado para encontrar los restantes se marca todas las celdas de la columna de Frecuencia Absoluta y se da F2, CONTROL-MAYUSCULA-ENTER y aparecen el resto de las frecuencias.

Sobre la misma hoja de Cálculo se construye la Tabla de Distribución de Frecuencias, la que se muestra en la Tabla 25.

Tabla 25. Tabla de distribución de frecuencias por intervalo para PR 8 talla de camarón.

VARIANTE 2 POR INTERVALO					
INTERVALO	LIM SUPERIOR	FREC ABS	FREC ACUM	FREC REL ABS	FREC REL ACUM
38 - 40	40	11	11	0.1833	0.1833
40 - 42	42	5	16	0.0833	0.2667
42 - 44	44	12	28	0.2000	0.4667

44 - 46	46	10	38	0.1667	0.6333
46 - 48	48	8	46	0.1333	0.7667
48 - 50	50	8	54	0.1333	0.9000
50 - 52	52	6	60	0.1000	1.0000

Fuente: Datos ficticios.

c) A partir de cualquiera de las dos tablas de Distribución de Frecuencias anteriores, y recordando que en la agrupación por intervalos los datos con igual valor de los límites, se coloca en el intervalo del límite superior, pueden realizarse el siguiente análisis:

- Si se suma las Frecuencias Relativas Absolutas en los intervalos desde 40 hasta 50 camarones por libra o restando en la Frecuencia Relativa Acumulada, la acumulada hasta 50 menos hasta 40 ($0.9000 - 0.1833 = 0.7177$), puede concluirse que un 71.77% de las libras de camarón de esa piscina cumple con la talla contratada, de 41 a 50 camarones por libra.
- Un 18.33% queda por debajo de la talla contratada (intervalo de 38 a 40)
- Un 10.00% está por encima de la talla contratada (intervalo de 50 a 52)
- Un 81.77% de la muestra tomada cumple o sobre cumple la talla contratada.

De lo anterior se deriva tomar decisiones por la Dirección de la Empresa sobre la extracción o no de camarones de esta piscina para la entrega al comprador.

1.6. Representación gráfica de las distribuciones de frecuencias

Las tablas de frecuencias y las representaciones gráficas son dos maneras **equivalentes** de presentar la información ya que ambas exponen de forma ordenada la información recogida en un estudio determinado. de manera general las formas gráficas resultan mucho más atractivas a la hora de brindar una determinada información.

En lo adelante se expondrán las alternativas más comunes de realizar estos tipos de representaciones gráficas. Los paquetes informáticos de estadística, presentan una amplia gama de representaciones gráficas, como anteriormente, se utilizará el EXCEL para la confección de las gráficas de la Distribuciones de Frecuencias

Diagramas de barras.

Los diagramas de barras se estructuran representando en uno de los ejes de coordenadas las clases o intervalos y en el otro eje de las frecuencias absolutas, acumuladas o frecuencias relativas.

En estos casos las alturas de las barras se hacen proporcionales a las frecuencias absolutas, acumuladas o relativas. Se utilizan fundamentalmente para las variables cualitativas, aunque se pueden aplicar también a variables cuantitativas discretas.

EJEMPLO 6. Para ejemplificar se tomarán los datos del ejemplo 2 relacionado con la satisfacción de los clientes en un hotel, cuya tabla 8 de Frecuencia aquí se repite:

Clases	Frecuencia absoluta	Frecuencia Absoluta Acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Muy satisfecho	4	4	0.200	0.200
Satisfecho	7	11	0.350	0.550
Moderadamente satisfecho	5	16	0.250	0.800
Insatisfecho	4	20	0.200	1.000
Total	20		1.0	

Para construir es gráfico de barras correspondiente a este ejemplo, se situarán en el eje de las abscisas las clases (Muy satisfecho, satisfecho, etc.) y en el eje de las ordenadas, los valores de frecuencia absoluta. Asociada a cada clase se levantará una columna a escala cuya altura coincida con la frecuencia absoluta de esa clase. El resultado final de muestra en el gráfico 1.

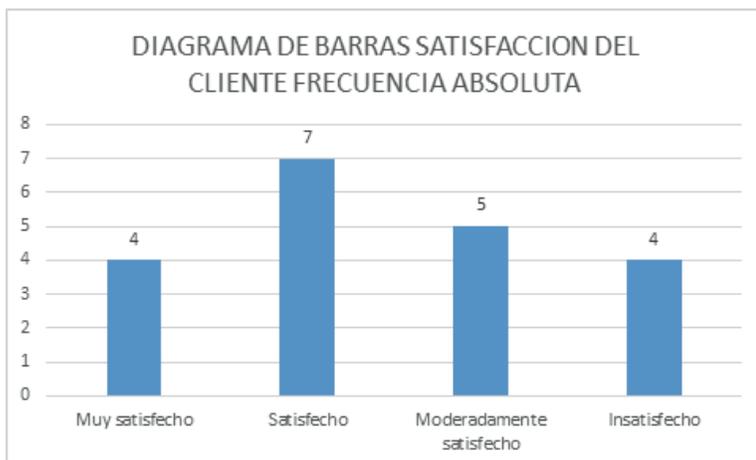


Gráfico 1. Diagrama de Barras del Ejemplo 6.

El resto de los gráficos de barras para las otras frecuencias se construye de forma similar, pero situando en el eje de las ordenadas los valores de las frecuencias acumuladas y relativas. Los gráficos correspondientes a este ejemplo se muestran en la Figura 28.

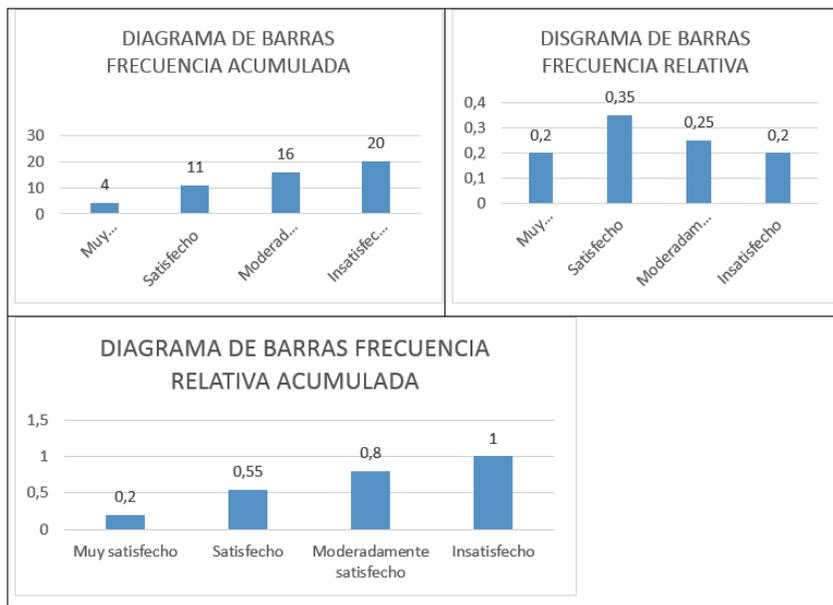


Figura 28. Gráficos de las frecuencias absoluta acumulada, relativa y relativa acumulada.

Aunque los gráficos anteriores se construyen fácilmente de manera manual, en la actualidad se utilizan los paquetes informáticos de estadística para construir los mismos. Los anteriores fueron realizados utilizando el EXCEL. Para ello se copia en una hoja EXCEL la Tabla de Distribuciones de Frecuencias, tal como se muestra en la Figura 29.

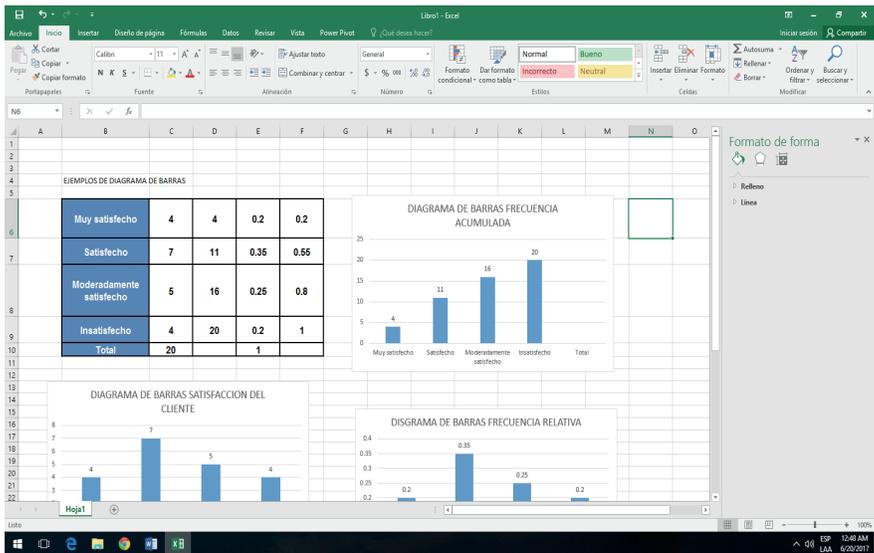


Figura 29. Hoja EXCEL para la construcción de Gráficos de Barras.

Como segundo paso se selecciona las columnas de las etiquetas y el de la frecuencia que se quiere representar, en este caso la Frecuencia Absoluta, se selecciona en la barra superior INSERTAR y dentro de los Gráficos Recomendados el de Barras o Columnas e inmediatamente aparecerá en la misma Hoja el gráfico que se muestra en la Figura 30.

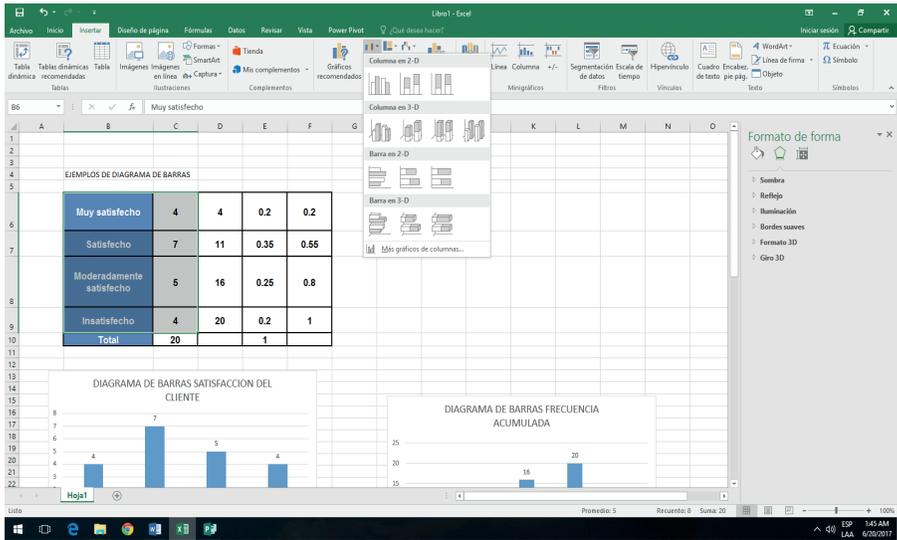


Figura 30. Hoja EXCEL para la construcción del Gráfico de Barras del Ejemplo 6.

Finalmente utilizando las herramientas del diseño del gráfico se adicionan las etiquetas y el título, lo que se muestra en la Figura 31.

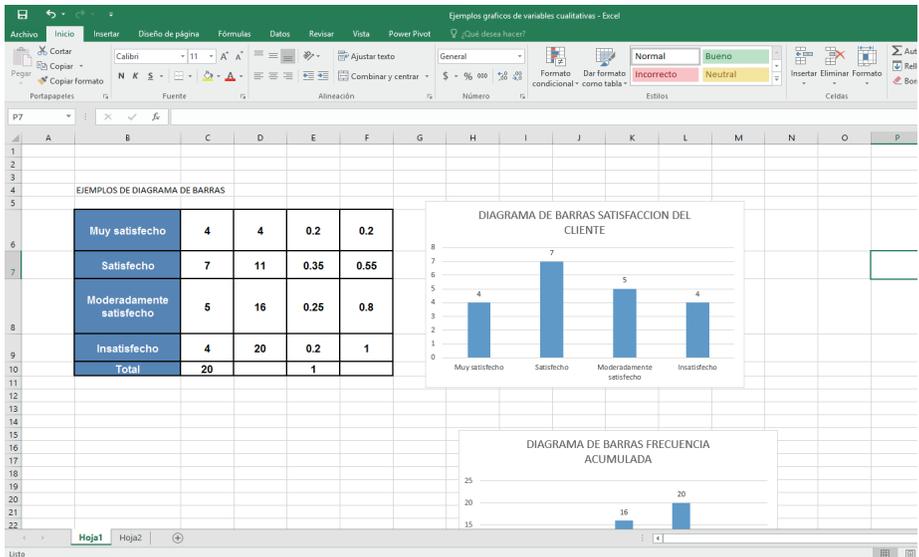


Figura 31. Gráfico de Barras de la Frecuencia Absoluta del Ejemplo 6.

Gráfico de sectores o de pastel

Otra alternativa gráfica para problemas de variables discretas que tiene el EXCEL es el denominado Diagrama de Sectores o de Torta o Pastel. Este representa dentro de una circunferencia, distintos sectores asociados a la frecuencia absoluta o relativa para cada clase. Para ello una vez que los datos han sido copiados en la Hoja EXCEL, se selecciona de la tabla las etiquetas de las clases y los datos de la frecuencia que se quiere graficar (absoluta o relativa) se busca en la barra superior INSERTAR y se selecciona el Gráfico por Sectores(circular, por anillo, de pastel) e inmediatamente aparecerá en la hoja el Gráfico por Sectores; después haga clic en uno de los sectores y seleccione adicionar etiquetas de datos y se tendrá el gráfico que se muestra en el gráfico 2.

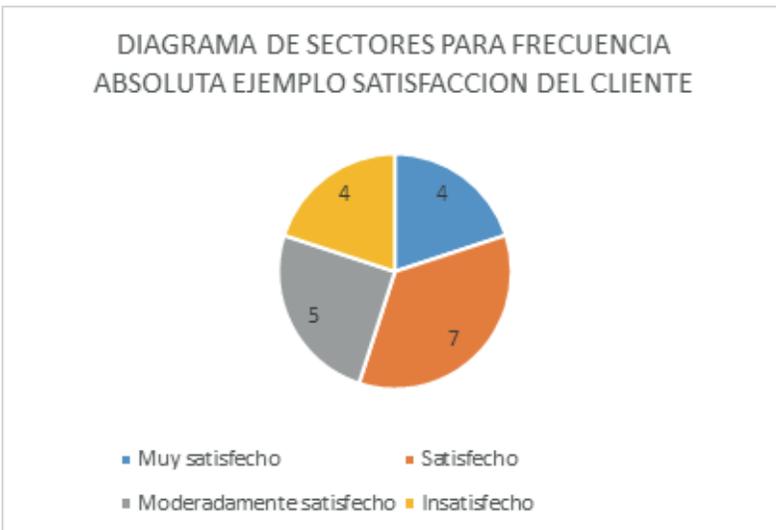


Gráfico 2. Sectores o pastel de la Frecuencia Absoluta.

Para el caso de la frecuencia relativa, se selecciona los datos de la Frecuencia relativa y siguiendo el mismo procedimiento se obtiene el gráfico 3.

DIAGRAMA DE SECTORES PARA FRECUENCIA RELATIVA EJEMPLO SATISFACCION DEL CLIENTE

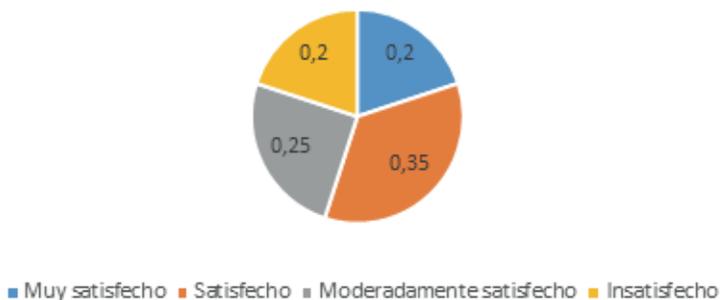


Gráfico 3. Sectores o pastel para Frecuencia relativa absoluta.

Histograma de frecuencia

Las representaciones gráficas en el caso de las variables numéricas o cuantitativas continuas, tienen por sus características, determinadas formas de expresión un tanto diferentes a las que se han utilizado para el caso de variables categóricas o cuantitativas discretas.

Como se estudio anteriormente, para el caso de las variables cuantitativas continuas, las frecuencias se acumulan dentro de intervalos numéricos. Esto implica que para la representación gráfica si se toma el eje de las abscisas para representar el valor de la variable bajo estudio, tiene que situarse los intervalos en los cuales se agrupan las frecuencias y para cada intervalo y en el eje de las ordenadas se situará los valores de la frecuencia. La representación gráfica en este caso, que se denomina Histograma, se construirá, dibujando una barra para cada intervalo, cuya altura será la frecuencia en ese intervalo.

EJEMPLO 7. Para ejemplificar se toman los datos del ejemplo 3 del tiempo de servicio en la caja de un banco.

INTERVALO	FRECUENCIA
[10 – 13] 5	
[13 – 16]	15
[16 – 19]	13

[19 – 22]	11
[22 – 25]	5
[25 – 28]	1
TOTAL	50

Para este caso en el eje de las abscisas se pondría los intervalos definidos y en el de las ordenadas el valor de las frecuencias. Para los dos primeros intervalos el histograma se muestra en el gráfico 4.

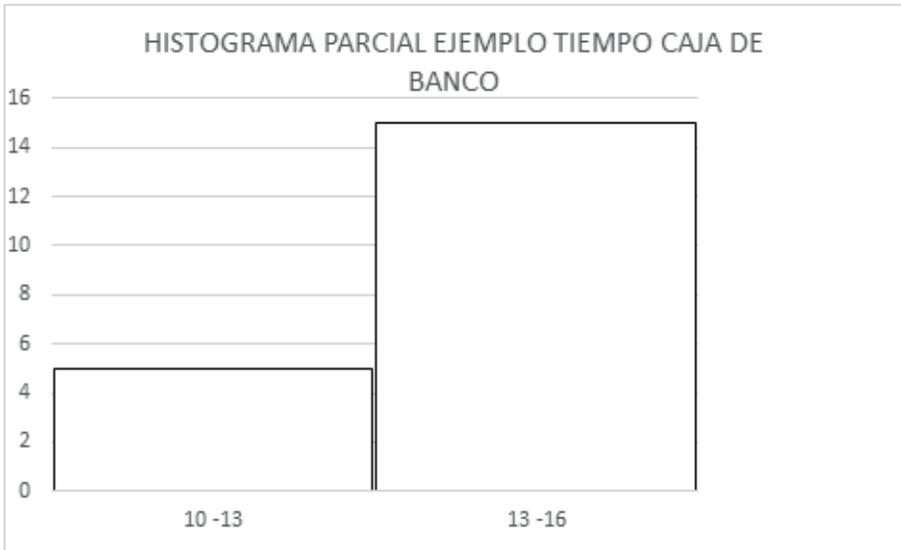


Gráfico 4. Histograma para una variable numérica continua.

Los Histogramas pueden construirse de forma directa utilizando el EXCEL. Para ello se colocan los datos en una columna de la Hoja EXCEL, en la barra superior se da INSERTAR y en la parte de los gráficos aparece el Histograma. Se marcan los datos y después se inserta el histograma, el cual el EXCEL construye de forma automática. Para ejemplificar se tomarán los datos del tiempo de servicio en la caja de un banco. Los datos en la hoja EXCEL quedarían como se muestra en la Figura 32.

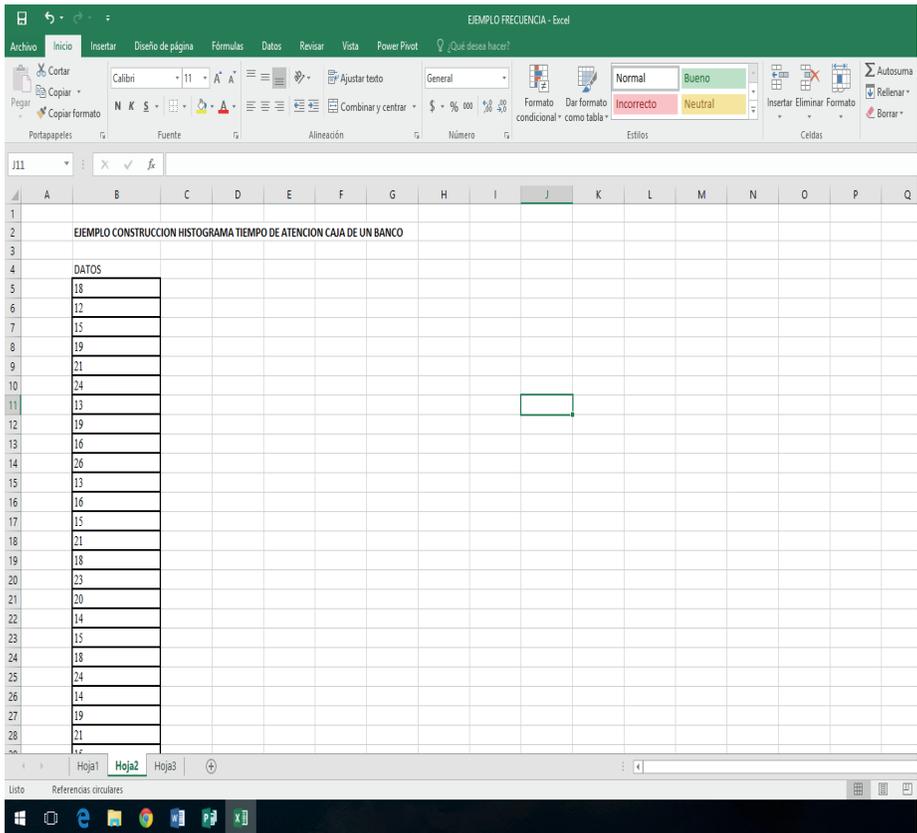


Figura 32. Hoja EXCEL con los datos para construir un Histograma.

Con la columna de los datos marcados, se va a INSERTAR y en los Gráficos se selecciona el Histograma, tal como se muestra en la Figura 33.

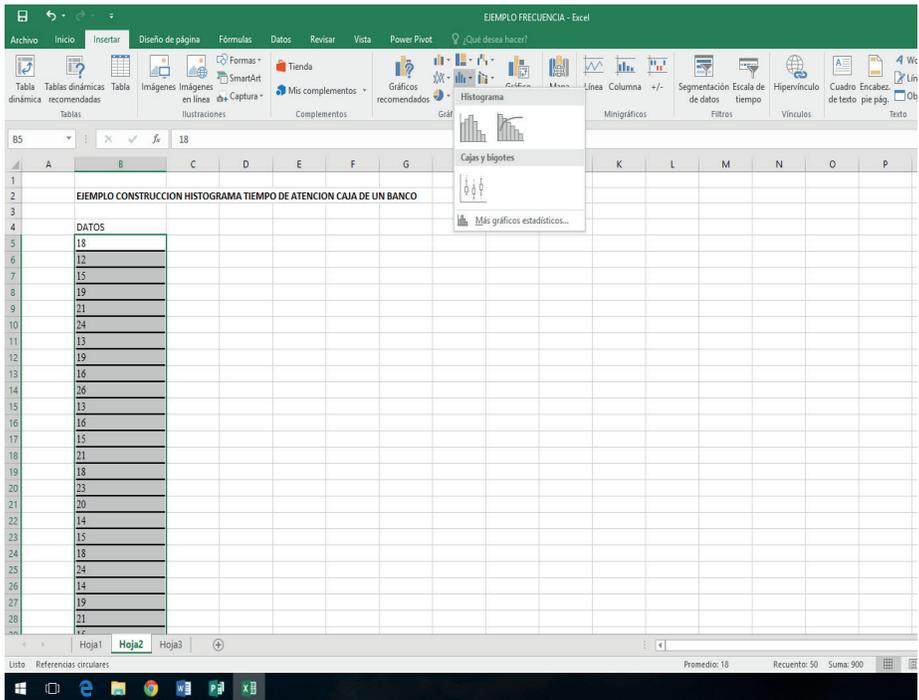


Figura 33: Construcción de Histograma utilizando EXCEL

Inmediatamente en la misma hoja EXCEL aparecerá el Histograma que se muestra en el gráfico 5.

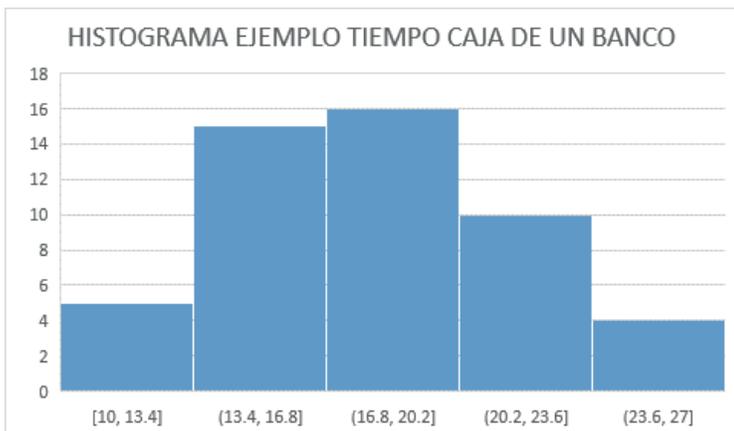


Gráfico 5. Histograma ejemplo 3 “tiempo de servicio en la caja de un banco”.

Note que se han seleccionado automáticamente 5 intervalos con una amplitud de 3.4 minutos cada uno de ellos. Utilizando las herramientas de edición de este gráfico se puede ajustar al ejemplo que se ha venido desarrollando con 6 intervalos y una amplitud de 3 minutos.

Para ello se da doble clic en el valor de los ejes y aparecerá una ventana, se marca opciones del eje y aparece una nueva ventana que permite hacer los ajustes deseados, cambiando la versión automática que trae por defecto y cambiando el número de intervalo y posteriormente con los mismos pasos la amplitud deseada para cada intervalo. El cambio del número de intervalos a 6 se muestra en la Figura 34.

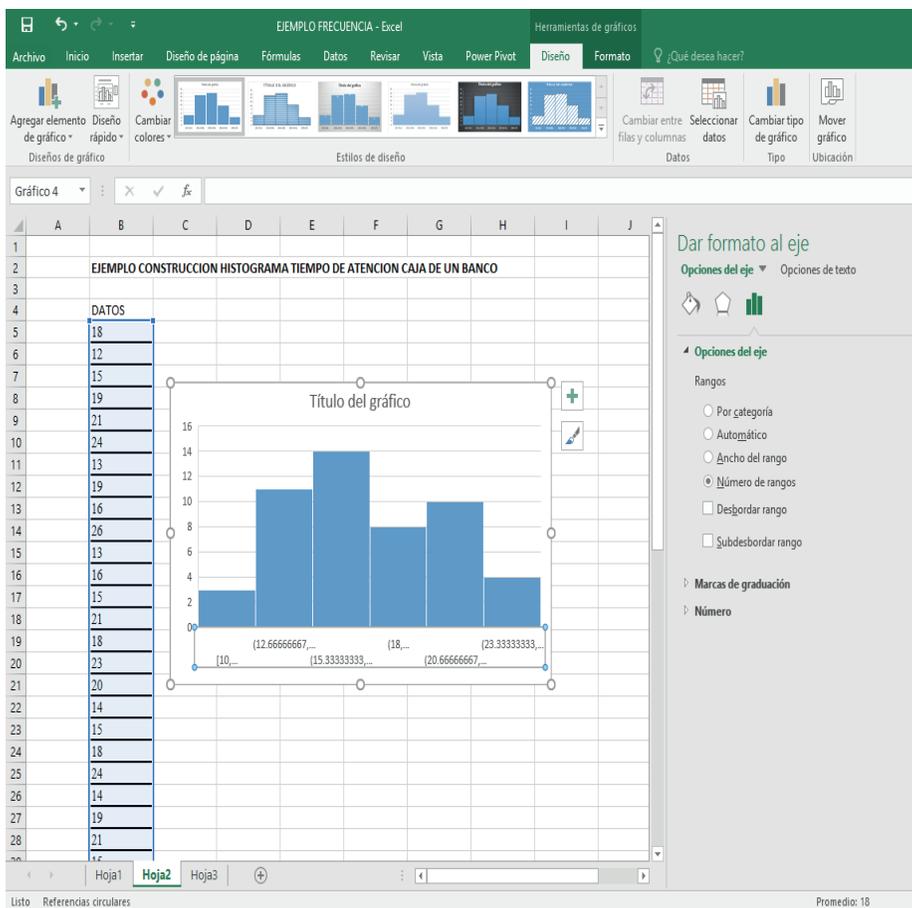


Figura 34. Edición del Histograma ejemplo 3 “tiempo de servicio en la caja de un banco”.

Finalmente cambiando la amplitud del intervalo a 3 minutos quedara el Histograma que se muestra en el gráfico 6.

HISTOGRAMA TIEMPO DE ATENCION CAJA DE UN BANCO

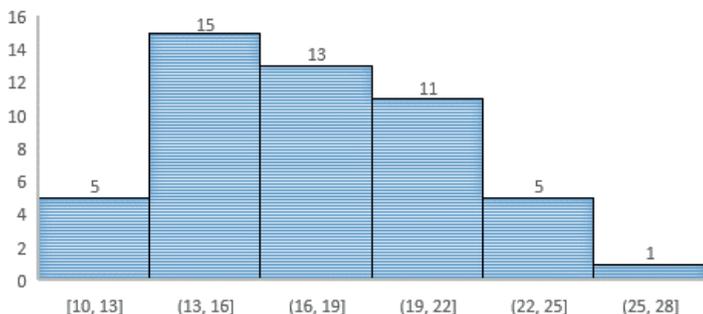


Gráfico 6. Histograma, Ejemplo 1.3 "tiempo de servicio en la caja de un banco" con 6 intervalos.

Este se corresponde con la tabla 6 de frecuencias absolutas construida anteriormente.

Polígonos de Frecuencias

Los Polígonos de Frecuencia, son otra forma de representación gráfica de las Distribuciones frecuenciales en variables discretas o continuas. Es bastante utilizado para representar las Distribuciones de frecuencia acumulada absoluta y relativa y permite conocer, en el caso de la Frecuencia relativa acumulada, la proporción o porcentaje que es menor o igual que un valor fijado de la variable.

Este se construye en el primer cuadrante de un sistema de coordenadas, situando en el eje de las abscisas el valor de la variable categórica o cualitativa o los límites superiores de cada intervalo definido y en la ordenada, la frecuencia absoluta o relativa de la variable categórica o el valor de la frecuencia en el intervalo para variables cuantitativas. Para cada par de valores quedará definido un punto en el cuadrante y al unir estos puntos por segmentos de rectas quedará representado el Polígono de frecuencias.

Estos gráficos son sencillos de construir utilizando el EXCEL e insertando un Gráfico de Línea con los datos descritos anteriormente para cada eje.

EJEMPLO 8. Para el caso del ejemplo 2 que es una variable cualitativa, la tabla 8 de Distribución de Frecuencia calculada fue:

Clases	Frecuencia absoluta	Frecuencia Absoluta Acumulada	Frecuencia relativa	Frecuencia relativa acumulada
Muy satisfecho	4	4	0.200	0.200
Satisfecho	7	11	0.350	0.550
Moderadamente satisfecho	5	16	0.250	0.800
Insatisfecho	4	20	0.200	1.000
Total	20		1.0	

Para este caso en el eje de abscisas se colocarán los valores de la variable, muy satisfecho, satisfecho, moderadamente satisfecho e insatisfecho y en el eje de las ordenadas los valores de la Frecuencia que se quiera graficar.

Si se quisiera obtener el Polígono de la Frecuencia Relativa Absoluta, se copiaría la Tabla anterior en una Hoja de Cálculo EXCEL, se marcan las columnas de los valores de la variable y la columna de la Frecuencia Relativa Absoluta, en la barra superior se escoge Insertar y se selecciona el Gráfico de Línea, tal como se muestra en la Figura 35

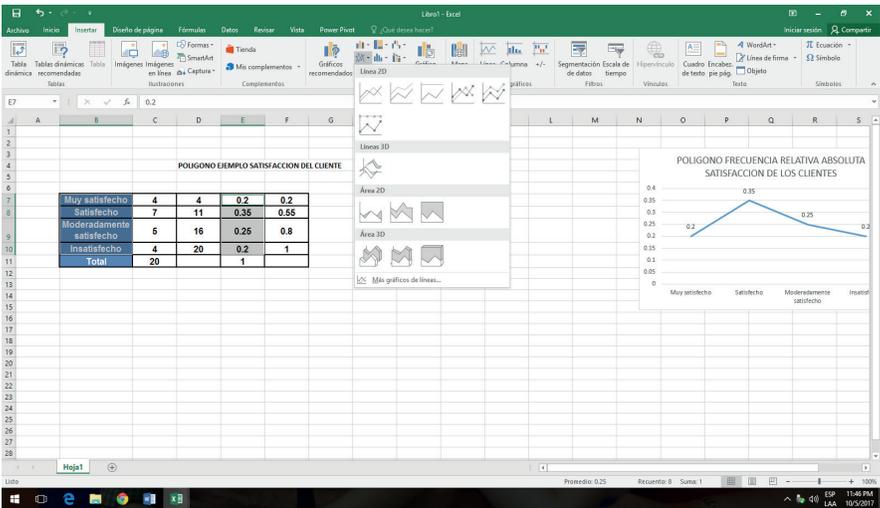


Figura 35. Ventana para la construcción de Polígonos de Frecuencia.

De forma inmediata aparecerá el Polígono de la Frecuencia Relativa Absoluta y utilizando las herramientas de diseño gráfico se completa el título y etiqueta de los datos, quedando como se muestra en el gráfico 7.

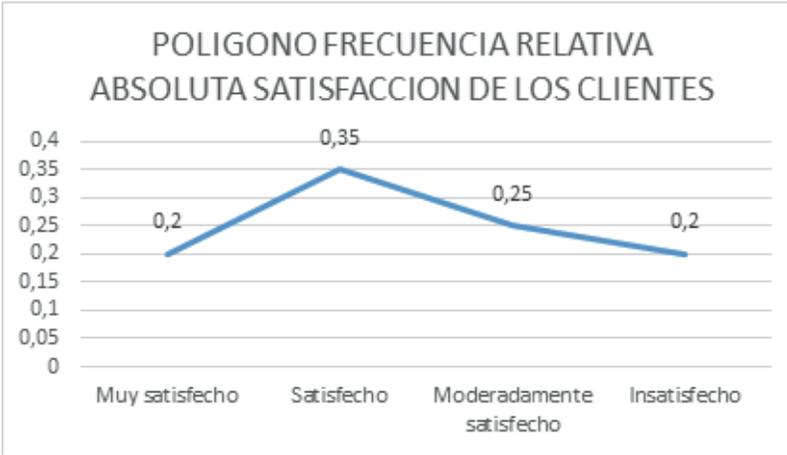


Gráfico 7. Polígono de Frecuencia Relativa Absoluta para Ejemplo 8.

El Polígono de Frecuencia Relativa Acumulada quedaría de la siguiente forma:

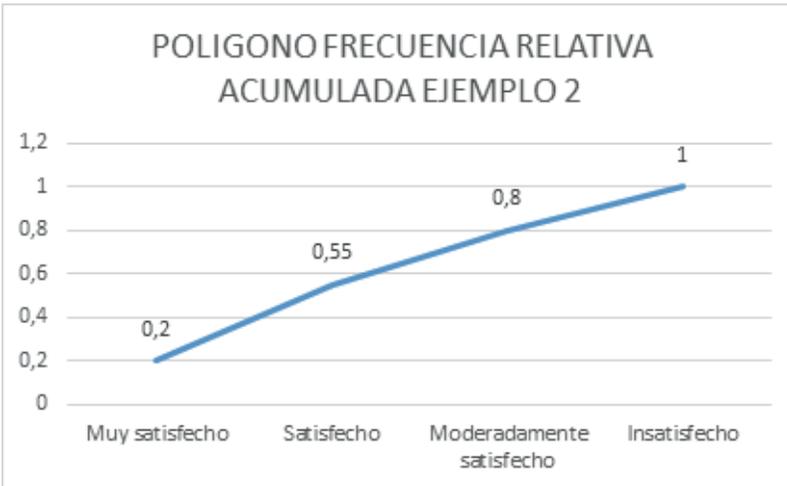


Gráfico 8. Polígono de Frecuencia Relativa Acumulada para Ejemplo 2.

EJEMPLO 9. Algo similar podría hacerse para el caso de las variables continuas, esto es:

1. En una Hoja EXCEL se copia la Tabla de Distribución de Frecuencias

- Se modifica la columna de los límites de intervalos, dejando solamente el límite superior y dando formato de texto a las celdas de esta columna.
- Se selecciona la columna de los límites superiores de los intervalos y cualquier otra columna de la Tabla de Distribución de Frecuencia.
- En la barra superior, se da Insertar y se elige el Gráfico de Líneas y aparece de forma inmediata el Polígono para la Frecuencia seleccionada. Utilizando las herramientas de diseño gráfico se completa con el título, las etiquetas de datos y cualquier otra información que se desee.

Para el ejemplo 3 del tiempo en la Caja de un Banco, la Hoja Excel, con los datos y los Polígonos de la Frecuencia Absoluta y de la Frecuencia Relativa Acumulada, se muestra en la siguiente Figura 36:

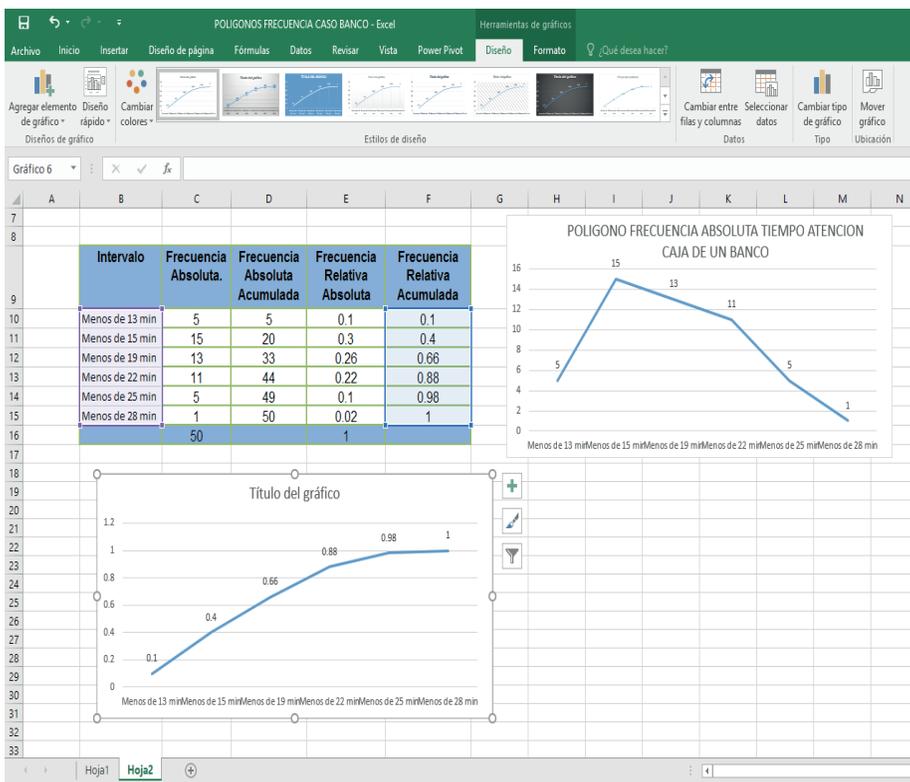


Figura 36. Polígonos de Frecuencia para una variable continua.

Del Polígono de Frecuencia Relativa Acumulada, puede observarse que un 66 % de los clientes que van al servicio de caja del banco están 19 minutos o menos y un 88%, 22 minutos o menos.

1.7. Problemas resueltos de construcción de gráficos

En esta sección se utilizarán los problemas resueltos anteriormente para el cálculo de las frecuencias y los gráficos se construirán directamente utilizando el EXCEL.

PR 9. Para el problema resuelto 1 (PR 1), relacionado con los defectos en la fabricación de muebles metálicos, construya:

- a) Diagrama de barras para la Frecuencia Absoluta.
- b) Diagrama de pastel para la Frecuencia Relativa Absoluta.
- c) Polígono de Frecuencia Relativa Acumulada

SOLUCION

Es preciso recordar que, para el caso de variables cualitativas, la construcción de gráficos de frecuencias, utilizando el EXCEL, se hace usando directamente la Tabla de Distribución de Frecuencias y copiándola en un Hoja EXCEL. Para este problema se muestra el resultado en la Figura 37.

TABLA DE DISTRIBUCION FRECUENCIAL PR 1.1 DEFECTOS EN MUEBLES METALICOS

DEFECTO	FRECUENCIA ABSOLUTA	FRECUENCIA ABSOLUTA ACUMULADA	FRECUENCIA RELATIVA	FRECUENCIA RELATIVA ACUMULADA
	f_i	F_i	f_{ri}	F_{ri}
ABOLLADURA	21	21	0.35	0.35
PINTURA	17	38	0.28	0.63
ENSAMBLAJE	12	50	0.2	0.83
CERRADURA	10	60	0.17	1
TOTAL	60		1	

Figura 37. Datos en Hoja Excel de la Tabla de Distribución Frecuencial del PR 1.

Para construir el Gráfico de Barras, se selecciona las columnas que se desean graficar, en este caso las etiquetas de las variables y la Frecuencia Absoluta, se escoge en la barra superior la función Insertar y en gráficos el de barras, tal como se muestra en la Figura 38.

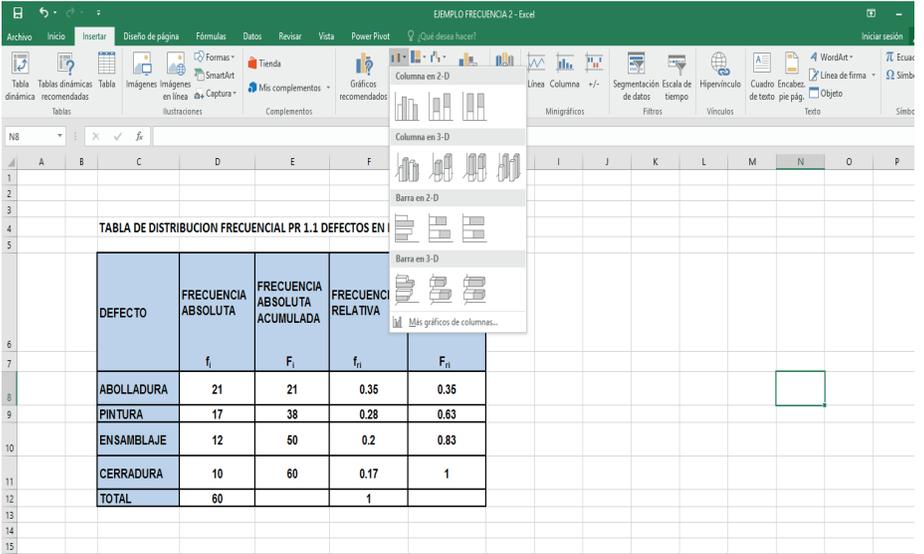


Figura 38. Ventana con los tipos de Gráficos de Barras de EXCEL.

Se elige el Gráfico de Barras en este caso e inmediatamente se muestra en la misma hoja el Gráfico deseado, lo que aparece en la Figura 39:

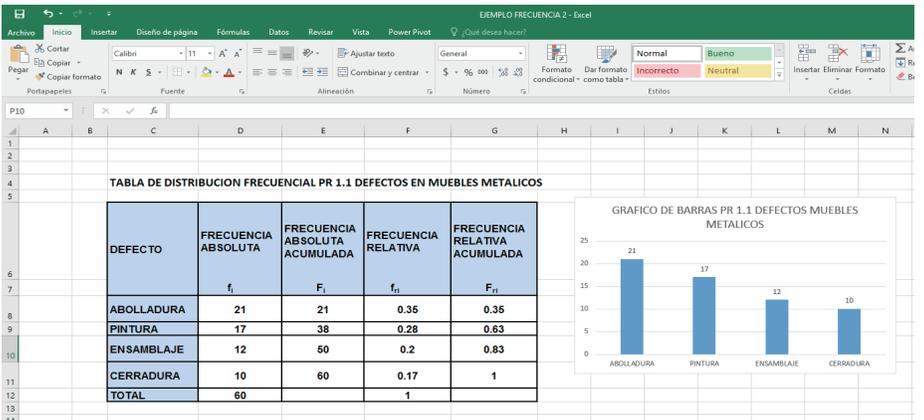


Figura 39. Gráfico de Barras para PR 1 defectos en muebles metálicos.

De forma similar puede obtenerse el Gráfico de Barras para las otras frecuencias que aparecen en la Tabla.

b) Para construir el Gráfico de Pastel se sigue el mismo procedimiento, pero en la elección del gráfico se escoge este y se mostraría el resultado que aparece en la Figura 40:

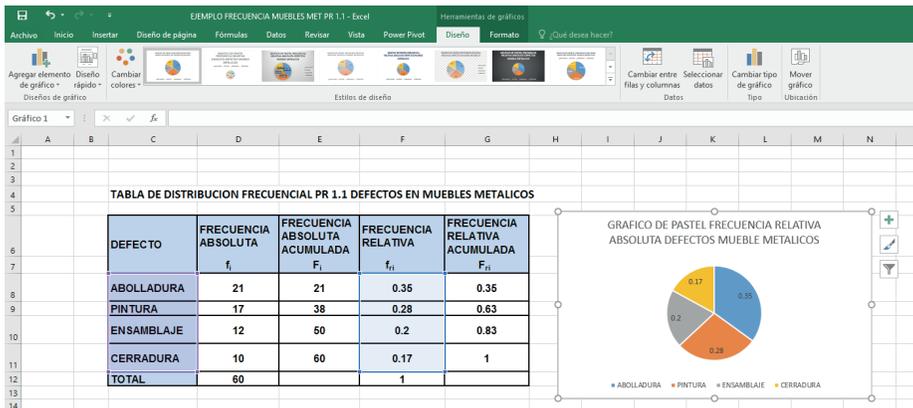


Figura 40. Diagrama de Pastel para PR 1 utilizando EXCEL.

c) Para construir el Polígono de Frecuencia Relativa Acumulado, en la misma hoja se marcan los datos de la variable y de la Frecuencia Relativa Acumulada, se da Insertar en la barra superior y se selecciona el Gráfico de Líneas y aparecerá en la hoja el gráfico y utilizando las herramientas de diseño se completa el título y las etiquetas de los datos. El resultado final se muestra en la Figura 41

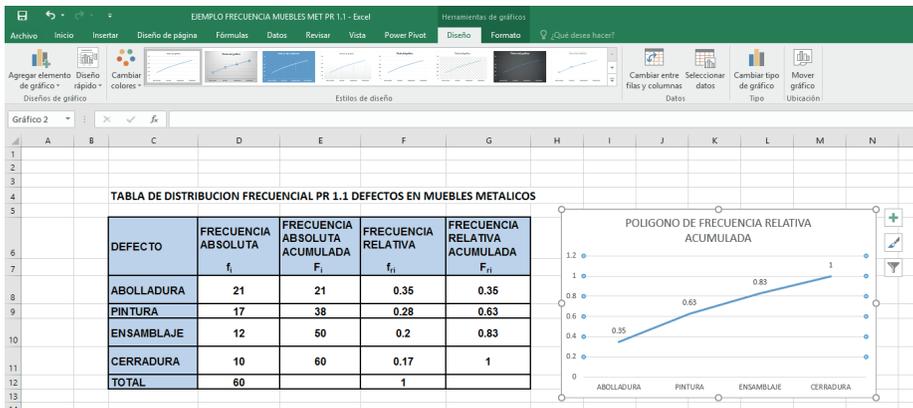


Figura 41. Polígono de Frecuencia Relativa Acumulada para PR 1.

PR 10. Basado en los datos del Problema Resuelto 4 (PR 4), relacionado al desecho de madera en una carpintería, construya:

- a) El Histograma para la Frecuencia Absoluta.
- b) El Polígono de Frecuencia Absoluta Acumulada y Relativa Acumulada

SOLUCION

a) La información sobre los kilogramos de desecho de madera por semana, que se mostró en el PR 4, es la siguiente:

26.3	32.5	23.00	25.2	19.3	21.6	34.0	32.6	30.5	20.8
28.3	35.1	24.5	21.3	27.0	30.1	18.3	21.9	24.4	23.6
22.3	23.0	27.8	32.6	34.4	23.2	29.0	31.4	26.5	28.9
24.6	27.3	30.2	32.5	19.0	23.6	30.0	24.6	27.7	28.0
29.0	21.5	22.4	20.8	27.8	31.1	25.7	22.8	28.6	27.2

Para construir los Histogramas utilizando el EXCEL, se debe colocar esta información en una columna de una hoja EXCEL, copiándola por fila o por columna de la tabla de datos anterior. Posteriormente con la columna de los datos marcada, se elige en la barra superior Insertar y se escoge entre los Gráficos el del Histograma, tal como se muestra en la Figura 42.

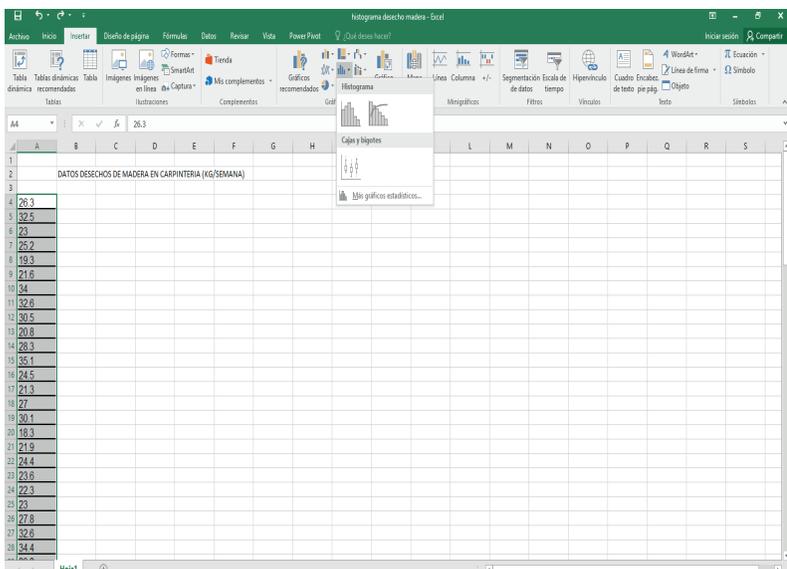


Figura 42. Datos del PR 4 y ventana de Histograma en hoja EXCEL.

Se selecciona el Histograma y de inmediato aparece la figura del mismo en la hoja de cálculo. Utilizando las herramientas de diseño de la figura se realizan los ajustes necesarios al Histograma: título, título de los ejes, cambio en la amplitud de los intervalos, etc. Ambos histogramas se muestran en la Figura 43.

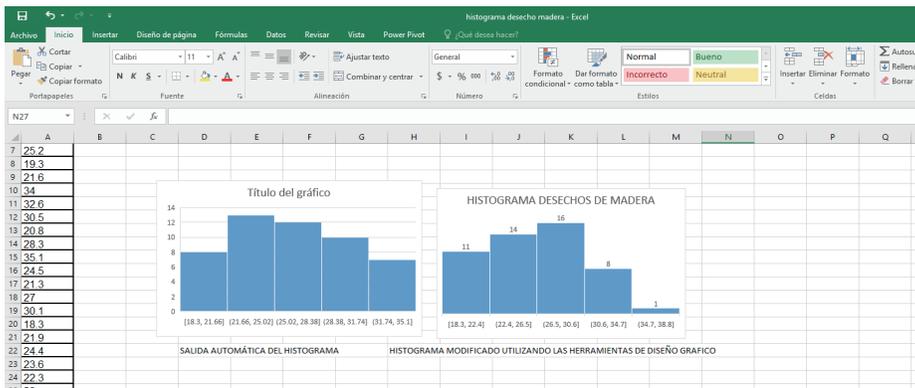


Figura 43. Histograma de Frecuencia Absoluta para PR 4 de desechos de madera.

En el Histograma modificado, se adicionó el título, las etiquetas de la frecuencia absoluta en cada intervalo y se fijó la amplitud del intervalo en un valor de 4.1 kg/semana, que fue el valor seleccionado en el PR 4.

b) Para construir los Polígonos de Frecuencias solicitados en este inciso, utilizando el EXCEL, se parte de la Tabla de Distribución de Frecuencias, obtenida en el PR 4, que fue la siguiente:

INTERVALOS PARA EL DESECHO EN KG/SEMANA	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	FRECUENCIA RELATIVA f_{ri}	FRECUENCIA RELATIVA ACUMULADA F_{ri}
18.3 - 21.7	8	8	0.16	0.16
21.7 - 25.1	13	21	0.26	0.42
25.1 - 28.5	12	33	0.24	0.66
28.5 - 31.9	10	43	0.20	0.86
31.9 - 35.3	7	50	0.14	1.00
TOTAL	50		1.00	

La tabla anterior se copia en una hoja EXCEL. Se modifica la columna de los intervalos, dejando solamente el límite superior del intervalo y en Formato de Celda se formatea estas, como celdas de texto.

Para obtener los Polígonos de Frecuencia, se selecciona en la Hoja EXCEL la columna de los intervalos y la otra columna que se desee graficar, por ejemplo, el de la Frecuencia Acumulada. Se escoge Insertar en la barra superior y se selecciona el gráfico de líneas e inmediatamente aparecerá en la misma hoja el Polígono de Frecuencia de la frecuencia acumulada. Esto se muestra en la Figura 44.

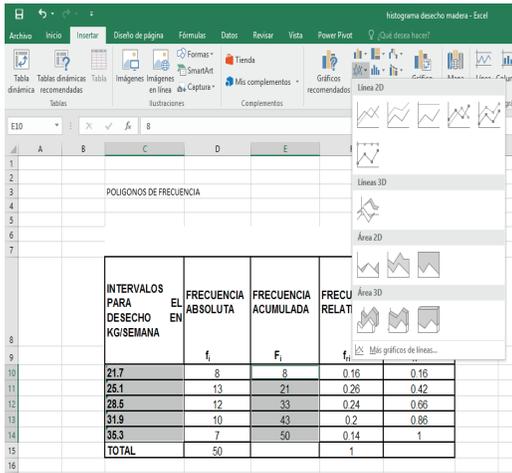


Figura 44. Hoja EXCEL para construir Polígonos de Frecuencias.

Sobre el gráfico obtenido y utilizando las herramientas de diseño gráfico, se adiciona la etiqueta de los datos, el título y cualquier otra modificación que se desee. Los gráficos finalmente quedarían de la siguiente forma:

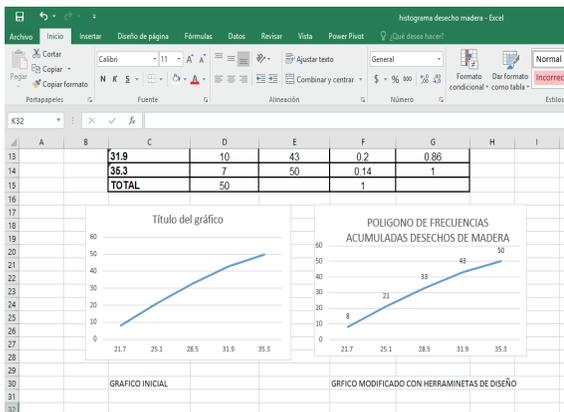


Figura 45. Hoja EXCEL con Polígonos de Frecuencias.

De forma similar se puede obtener el Polígono para la Frecuencia Relativa Acumulada, seleccionando la columna de intervalo y de la Frecuencia Relativa Acumulada en la Tabla de los datos y siguiendo los mismos pasos se obtiene la siguiente figura 46.

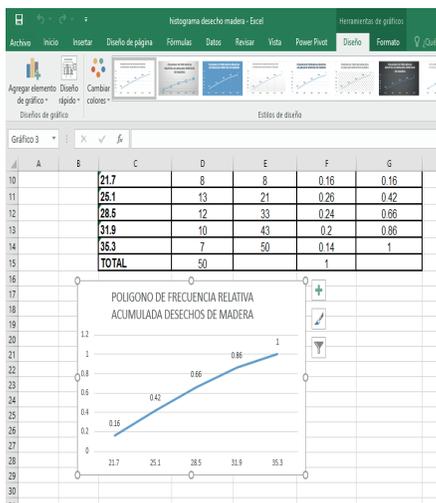


Figura 46. Hoja EXCEL con Polígono de Frecuencia Relativa.

PR 11. Con la información dada en el Problema Resuelto 7 (PR 7):

- El Diagrama de Barras para la Frecuencia Absoluta.
- El Gráfico de Pastel para la Frecuencia Relativa Absoluta
- El Polígono de Frecuencia Relativa Acumulada

SOLUCION

En este problema, que se refería a la calidad del servicio de una Cooperativa de transporte de pasajeros por ómnibus interprovincial, la variable es cualitativa y por tanto las representaciones gráficas solicitadas, utilizando el EXCEL, se construirán directamente de la Tabla de Distribución de Frecuencias, la cual fue encontrada para este problema y es:

VARIABLE	FREC ABSOL	FREC ACUM	FREC REL ABS	FRS REL ACUM
E	5	5	0.08	0.08
MB	11	16	0.18	0.27

B	15	31	0.25	0.52
R	16	47	0.27	0.78
M	8	55	0.13	0.92
P	5	60	0.08	1.00

Con esta información en una Hoja de Cálculo EXCEL, se procede a encontrar los gráficos solicitados.

a) Gráfico de Barras para la Frecuencia Absoluta.

Se marca los valores de las dos primeras columnas, esto es la categoría de la variable y la Frecuencia Absoluta y se va a la Barra Superior y se selecciona Insertar - Gráfico de Barras e inmediatamente en la misma hoja aparecerá el Gráfico de Barras correspondiente, tal como se muestra en la figura:

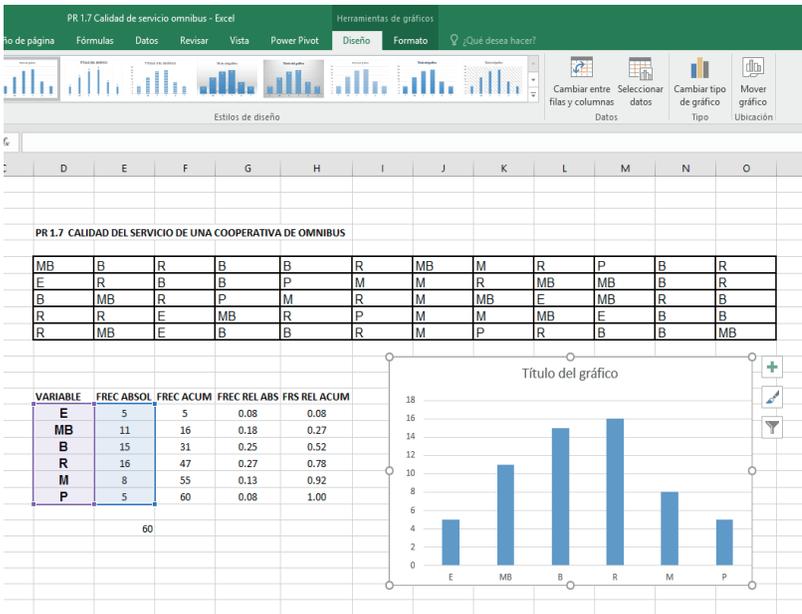


Figura 47. Hoja ESCEL Diagrama de Barras PR 7.

Finalmente, con el editor de gráficos, se completa el título y las etiquetas de los datos:

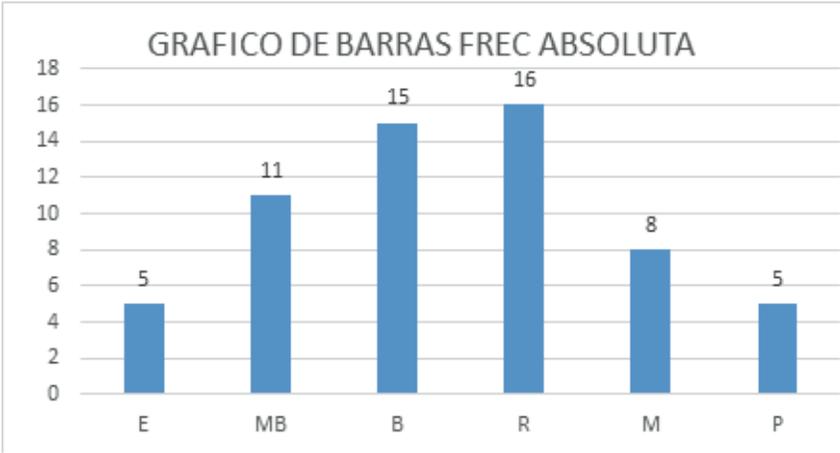


Gráfico 9. Gráfico de Barras editado con EXCEL de la Frecuencia Absoluta del PR 7.

b) Para la construcción del Gráfico de Pastel de la Frecuencia Relativa Absoluta se procede de forma similar, pero se selecciona el gráfico de pastel y este queda como:

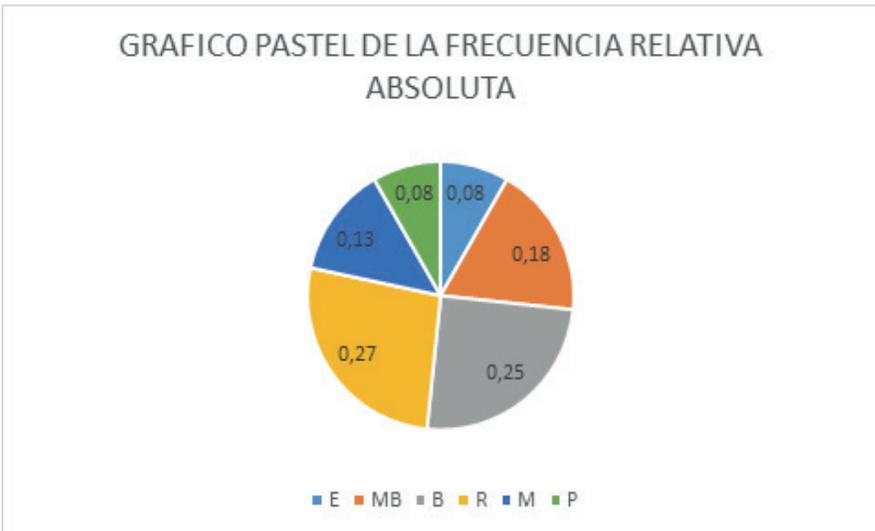


Gráfico 10. Gráfico de Pastel con EXCEL para Frecuencia Relativa PR 7.

c) El Polígono de Frecuencia Relativa Acumulada se construye seleccionando las columnas de las categorías y la de la Frecuencia Relativa Acumulada y se inserta un Gráfico de Líneas. Este se muestra en el gráfico 11.

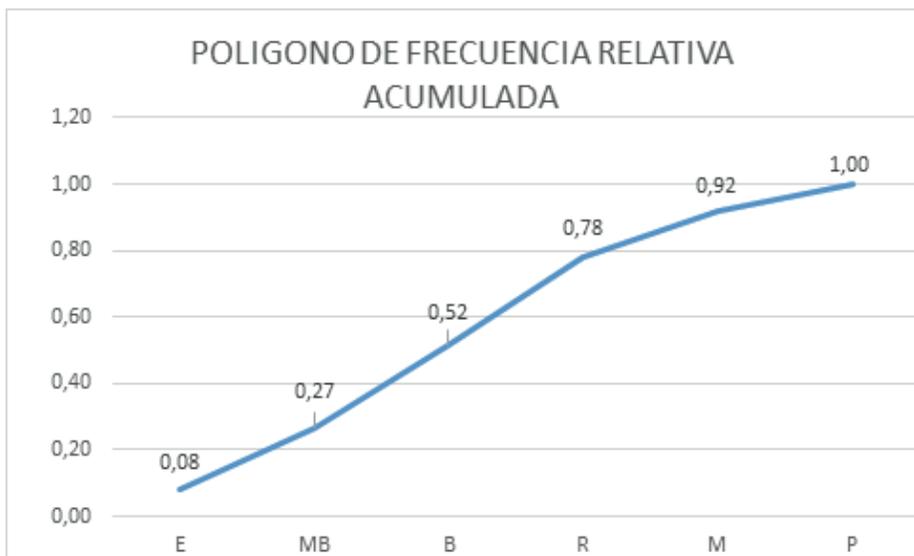


Gráfico 11. Polígono de Frecuencia Relativa acumulada construida con EXCEL del PR 7.

A partir de los gráficos pueden realizarse análisis similares a los expuestos en el PR 7.

PR 12. Utilizando los datos del problema resuelto 8 (PR 8), en su organización por intervalos, construya utilizando EXCEL:

- El Histograma de Frecuencia Absoluta.
- El Polígono de Frecuencia Relativa Absoluta.
- El Polígono de Frecuencia Relativa Acumulada.

SOLUCION

Este problema se resolverá de forma completa, esto es, primero se obtendrá la Tabla de Distribuciones de Frecuencias y después los gráficos solicitados en este problema.

Debe recordarse que: para construir el Histograma, los datos del problema deben colocarse en una misma columna de la Hoja EXCEL. Se incorpora en la Hoja los límites de los intervalos y la columna de la Frecuencia Absoluta y en la celda asociada al primer intervalo en esta columna se marca y se selecciona la función FRECUENCIA y aparece la ventana de esa función que

se llena con la información solicitada: en datos se pone la columna de los datos y en rango la columna de los límites superiores. La Hoja resultante se muestra en la Figura 48.

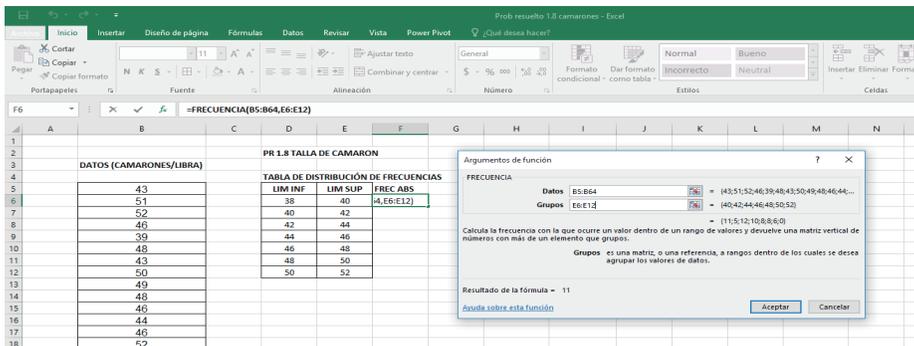


Figura 48. Construcción de la Tabla de Frecuencia Absoluta del PR 8 utilizando el EXCEL.

Dando Aceptar aparece la Frecuencia Absoluta asociada al primer intervalo y para hallar el resto, se marca toda la columna de la Frecuencia y se da F2+CONTROL-MAYUSCULA-ENTER y aparecen el resto de las Frecuencias Absolutas. Estas y la Tabla completada de la Distribución de Frecuencias se muestra en la Figura 49.

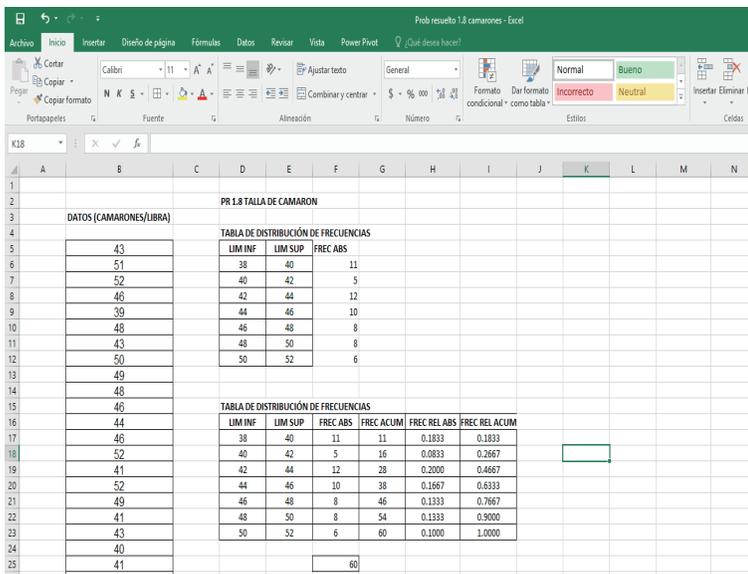


Figura 49. Tabla de Distribución de Frecuencias construida con EXCEL para PR 8.

a) Para construir el Histograma de Frecuencia Absoluta, se marca la columna de los datos y en Insertar se selecciona el gráfico del Histograma y aparece un histograma para este problema, pero en este caso automáticamente ha seleccionado solo 4 intervalos. Para que el histograma tenga los 7 intervalos que se vienen trabajando se da un doble “clic” en el valor de los intervalos y aparece una ventana para modificar el diseño, tal como se muestra en la Figura 50.

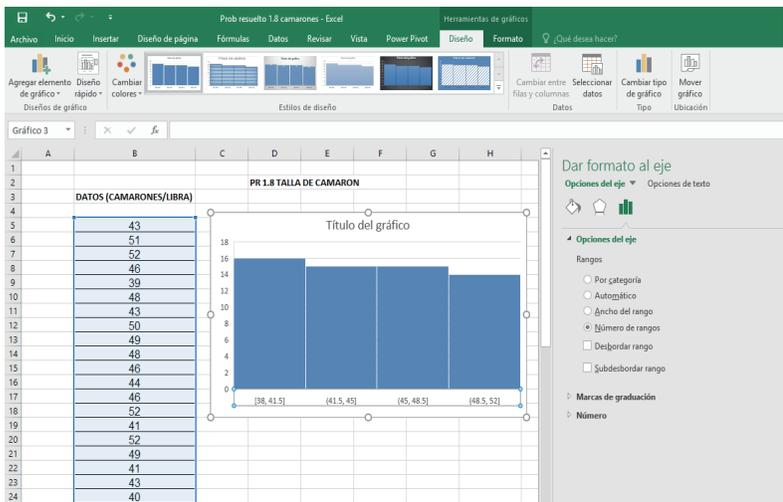


Figura 50. Histograma de Frecuencia Absoluta utilizando EXCEL para PR 8.

Seleccionando el Número de rangos como 7 de inmediato el Histograma aparece con los 7 intervalos deseados, el que se muestra en la Figura 51.

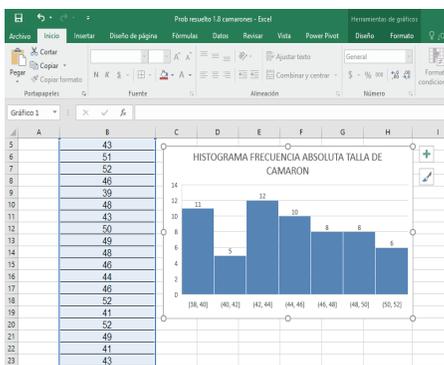


Figura 51. Histograma de Frecuencia Absoluta para PR 8 utilizando EXCEL.

b) Para construir el Polígono de Frecuencia Relativa Absoluta y de Frecuencia Relativa Acumulada, se utiliza la Tabla de Distribución de Frecuencias ya calculada. Para ello primeramente se le da formato de texto a la columna que tiene los límites superiores agregando alguna letra o palabra, se marca esa columna y la de la Frecuencia Relativa Absoluta y se da Insertar, seleccionando el Gráfico de Líneas y aparecerá el Polígono correspondiente. Posteriormente utilizando las herramientas de diseño gráfico se ponen las etiquetas de los datos y el título quedando finalmente como se muestra en la Figura 52.

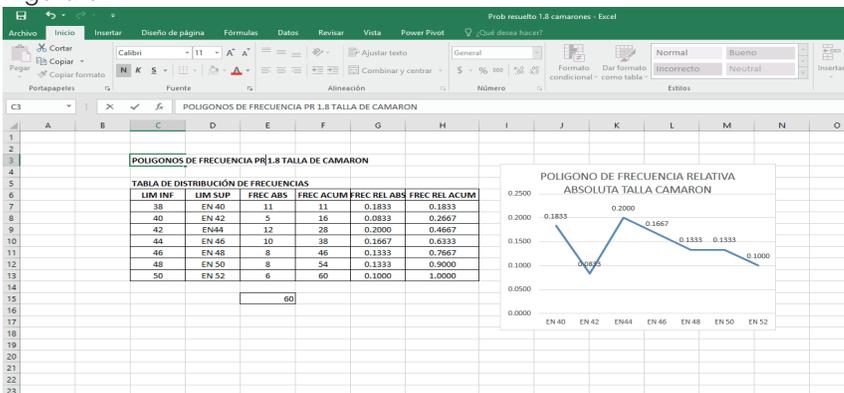


Figura 52. Polígono de Frecuencia Relativa para PR 8 utilizando EXCEL.

c) De forma similar, para hallar el Polígono de Frecuencia Relativa Acumulada, se cambia el texto en la 2^{da} columna y se marcan la columna Limite Superior y Frecuencia Relativa Acumulada, se escoge Insertar, Gráfico de Línea y saldrá el Polígono correspondiente y con el diseño gráfico se le da la terminación al título y etiqueta de los datos. Este se muestra en la Figura 53.

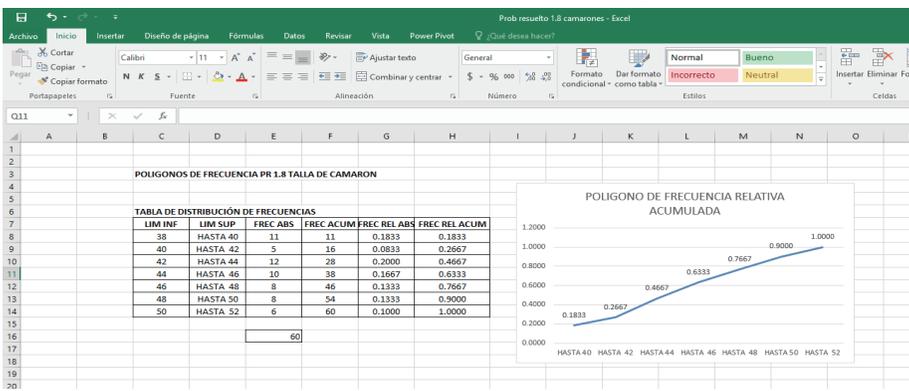


Figura 53. Polígono de Frecuencia Acumulada del PR 8 utilizando EXCEL.

1.8. Problemas propuestos de distribución de frecuencias y su representación gráfica

PP1. Una empresa metalúrgica está realizando una investigación sobre el proceso de planificación estratégica en la misma. Para ello han realizado una encuesta a 80 trabajadores, de ellos 30 que ocupan responsabilidades técnicas y el resto obreros directos de la producción. A una pregunta sobre la participación en la elaboración de los objetivos estratégicos, los encuestados han respondido de la siguiente forma:

CS	AV	AV	CN	N	N	CN	AV	AV	CN
N	CN	AV	AV	N	N	CN	AV	CN	N
CN	AV	AV	CN	N	CN	N	CN	AV	CN
CS	AV	CN	N	S	CS	CS	CN	CN	AV
CS	CN	N	N	CN	S	CS	N	CN	CN
CN	CS	CS	AV	AV	CS	S	AV	CS	AV
AV	S	CN	CS	CS	AV	S	CS	S	CS
CS	S	AV	CS	AV	CN	CS	CS	S	AV

Leyenda:

S: Siempre

CS: Casi siempre

AV: A veces

CN: Casi nunca

N: Nunca

Las tres últimas filas corresponden a las respuestas que dieron los 30 técnicos seleccionados para la encuesta.

- Construya la Tabla de Distribución Frecuencias para todos los encuestados.
- Construya la Tabla de Distribución de Frecuencias para los obreros y para los técnicos por separado.
- Construya un Gráfico de barras para cada uno de los tipos de encuestados.
- Analice en base a la información anterior como se puede valorar la participación de los trabajadores y la de los técnicos en la elaboración de los objetivos estratégicos.

PP 2. Una empresa que produce pulpa para jugos naturales, puede elaborar

diariamente una cantidad aleatoria de pulpa. Se ha tomado una muestra de 60 días de producción, para realizar un análisis del comportamiento productivo, en toneladas por día de pulpa, y los datos se muestran en la siguiente tabla:

162	156	172	165	163	180	159	160	154	166
176	181	169	182	176	179	180	168	162	178
180	179	168	164	176	170	182	176	178	169
178	175	184	180	179	169	175	175	181	169
176	179	182	176	168	163	176	183	179	171
181	178	165	169	176	181	179	182	168	177

- Construya la tabla de Distribución de Frecuencias para este problema.
- Construya el Histograma de Frecuencia.
- Construya el Polígono de Frecuencia Relativa Acumulada
- Analice los resultados.

PP 3. En un Hotel los ingresos semanales en miles de dólares es una cantidad aleatoria. Se ha tomado una muestra de las últimas 50 semanas y los datos se muestran en la siguiente tabla:

80.32	75.36	81.40	68.50	78.35	69.23	80.24	77.46	81.36	69.34
75.56	65.60	81.25	69.45	70.28	78.50	74.42	80.56	79.34	68.25
82.58	64.56	76.75	69.00	73.28	76.70	80.34	69.90	81.15	78.65
80.45	76.38	77.00	80.45	78.40	82.40	76.40	72.45	80.24	78.00
67.90	76.60	73.50	80.50	76.55	69.90	71.00	76.80	80.60	75.70

- Calcule la Tabla de Distribución de Frecuencias para este problema.
- Represente el Histograma de la Frecuencia Absoluta.
- Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- Analice los resultados.

PP4. Una empresa de jugos naturales, va a lanzar un nuevo producto al mercado en envases de 500 ml y desea saber qué tipo de envases prefiere el consumidor. Para ello ha realizado una encuesta a 80 consumidores potenciales donde le informa el tipo de envase y la relación con el precio de venta y la duración del producto, para que seleccione el tipo de envase que considera más adecuado. Los tipos propuestos de envase son: vidrio (VD),

plástico (PL), aluminio (AL) o tetra pak (TP). Los resultados de la encuesta se muestran en la siguiente tabla:

PL	AL	AL	VD	TP	PL	PL	TP	AL	VD
AL	TP	PL	VD	PL	PL	AL	TP	VD	AL
AL	PL	VD	PL	PL	TP	VD	TP	AL	PL
PL	AL	TP	TP	PL	PL	AL	PL	VD	PL
TP	PL	AL	AL	PL	VD	PL	TP	VD	PL
VD	VD	PL	AL	PL	TP	PL	AL	VD	PL
VD	TP	VD	PL	AL	PL	VD	PL	PL	VD
TP	AL	AL	PL	VD	PL	PL	VD	AL	TP

- Construya la Tabla de Frecuencias para este problema
- Represéntela gráficamente por un Diagrama de Barras o un Diagrama de Pastel.
- Realice un análisis de los resultados de la encuesta.

PP 5. Una planta procesadora para la limpieza del arroz, produce diariamente una cantidad aleatoria de toneladas de arroz limpio. Se ha tomado una muestra de 40 días de trabajo y la producción diaria en toneladas se muestra en la siguiente tabla:

46.2	42.7	38.4	39.0	40.4	41.6	44.3	42.0	38.7	40.3
45.1	40.5	35.9	39.5	44.2	42.6	37.8	40.3	41.3	38.2
39.6	45.0	44.2	39.8	37.4	40.3	40.0	41.6	38.5	40.8
44.3	36.8	38.2	45.1	44.6	41.5	38.3	39.4	40.6	38.1

- Construya la Tabla de Distribución de Frecuencias para este problema.
- Represente el Histograma de la Frecuencia Absoluta.
- Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- Analice los resultados.

PP 6. Se desea analizar los resultados de la evaluación de los trabajadores de una gran empresa y para ello se ha tomado una muestra de 50 trabajadores. Estos pueden ser evaluados en las siguientes categorías: Excelente (E); Muy Satisfactorio (MS); Satisfactorio (S); Regular (R) y Mal (M). En la siguiente tabla se muestra los resultados de la evaluación de los 50 trabajadores seleccionados:

MS	S	S	M	R	S	E	MS	R	S
E	M	S	S	R	M	E	S	S	MS
M	S	MS	R	S	R	MS	S	S	R
R	MS	E	S	R	MS	R	M	E	S
E	R	R	MS	S	S	R	S	R	R

- Construya la Tabla de Distribución de Frecuencias para este problema
- Representéla gráficamente por un Diagrama de Barras y con un Diagrama de Pastel.
- Realice un análisis de los resultados de la evaluación en esa empresa.

PP 7. El consumo de energía eléctrica, en kW-h, durante un mes en una pequeña empresa metalúrgica, es aleatoria. Se ha tomado una muestra de ese consumo en los últimos 36 meses y los valores se muestran en la siguiente tabla:

1560	1420	1542	1489	1501	1468	1567	1503	1498	1500
1456	1502	1523	1478	1490	1508	1479	1506	1488	1516
1488	1509	1516	1493	1501	1485	1499	1504	1523	1498
1497	1531	1511	1489	1532	1517				

- Construya la Tabla de Distribución de Frecuencias para este problema.
- Represente el Histograma de la Frecuencia Absoluta.
- Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- Analice los resultados del consumo de energía eléctrica de esta empresa.

PP 8. La cantidad de litros de combustible que consume cada semana, una pequeña empresa transportista de carga, es una cantidad aleatoria. Se desea analizar este consumo y para ello se ha tomado una muestra de las últimas 40 semanas y los valores se muestran en la siguiente tabla:

624	590	609	589	590	607	634	580	578	623
600	588	604	575	604	577	588	607	618	590
617	598	609	621	582	591	616	600	581	596
625	605	598	587	612	600	593	597	610	601

- Construya la Tabla de Distribución de Frecuencias para este problema.

- b) Represente el Histograma de la Frecuencia Absoluta.
- c) Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- c) Analice los resultados del consumo de combustible de esta empresa.

PP 9. Una Empresa Cafetalera realiza la cosecha de café oro. En un muestreo de las hectáreas ya cosechadas se ha recogido la información del rendimiento en quintales por hectárea y estas se muestran en la siguiente tabla:

82.5	87.2	80.4	76.3	83.8	78.9	77.5	81.8	82.0	83.6
88.0	79.7	84.6	83.7	80,8	78.0	81,4	83,9	87.0	76.7
76.9	80.4	84.8	85.0	83.2	78.6	79.0	81.7	84,1	88.2
77.4	78.9	84.3	86.1	80.3	79.0	85,6	82,8	80.0	79.3

- a) Construya la Tabla de Distribución de Frecuencias para este problema.
- b) Represente el Histograma de la Frecuencia Absoluta.
- c) Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- c) Analice los resultados del rendimiento por hectárea en esa empresa.

PP 10. Una empresa produce mini dosis de azúcar refino en paquetes de 5 gramos. Han llegado algunas quejas de los clientes de que los paquetes no cumplen con está especificación. Se ha tomado una muestra de 50 paquetes de la producción de un día y los resultados se muestran a continuación:

4.75	4.89	5.09	5.00	4.85	5.03	5.04	4.98	4.91	5.03
4.89	5.06	4.95	5.08	5.01	5.00	4.93	4.95	5.01	4.87
5.01	4.98	5.03	5.02	4.93	4.97	4.91	5.06	5.05	4.99
4.98	4.89	5,03	5.06	4.98	4.95	4.98	5.00	5.03	4.94
4.97	5.07	4.99	5.01	5.02	4.94	4.96	5.01	5.00	4.97

- a) Construya la Tabla de Distribución de Frecuencias para este problema.
- b) Represente el Histograma de la Frecuencia Absoluta.
- c) Construya el Polígono de Frecuencia Absoluta y Relativa Acumulada.
- c) Analice los resultados del rendimiento por hectárea en esa empresa.

Capítulo II. Medidas descriptivas

2.1. Estadígrafos de tendencia central

Un aspecto importante dentro de la Estadística Descriptiva es la de ofrecer algunas medidas de los datos que permita identificar características esenciales de los mismos. Un ejemplo de ello es el denominado promedio, muy utilizado en la vida práctica y del cual se oye hablar desde tempranas edades, peso promedio, temperatura promedio, salario promedio y que intuitivamente se sabe que es un valor que sintetiza en un único número, el comportamiento de una variable que toma valores aleatorios.

Para poder estudiar las medidas descriptivas es necesario dejar claro los conceptos de la Estadística Descriptiva es importante destacar los conceptos de parámetros y estadígrafos:

Parámetros; son cantidades numéricas calculadas sobre una población. A partir de este concepto en Estadística generalmente se busca el cálculo de determinados parámetros de tal manera que ello permita resumir en unos pocos números las características de una población.

Estadígrafos; son cantidades numéricas calculadas a partir de muestras de una población. En función de los estadígrafos se puede resolver el problema de hacer determinaciones aproximadas de los parámetros de una población, cuando por las características de la misma, resulte imposible abarcarla completamente. En tales casos, los estadígrafos son denominados también como “estimadores”.

Los estadígrafos que se estudiarán estarán en función de lograr resumir las características fundamentales en muestras seleccionadas, a partir de poblaciones son: de tendencia central, de dispersión, de posición y de forma.

Son estadígrafos que indican aquellos valores con respecto a los cuales los datos parecen agruparse alrededor de un valor ubicado en el centro del conjunto de los datos. De ellos, se estudiarán: media aritmética o promedio, trimedia, mediana y moda, que son las tres medidas de tendencia central de mayor uso, con excepción de la trimedia, en Estadística Descriptiva.

2.1.1. Media aritmética o Promedio

Sea una muestra de tamaño “n” tomada desde una población, cuyos valores son , entonces, la Media Aritmética o Promedio, que se denota como , se calcula por:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

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

Si los “n” datos estuvieran agrupados en “m” clases o intervalos y y_i es la Frecuencia Absoluta en el intervalo i -ésimo, entonces una forma abreviada y aproximada de calcular la Media Aritmética o Promedio es:

$$\bar{x} = (f_1 * y_1 + f_2 * y_2 + \dots + f_m * y_m) / n$$

$$\bar{x} = \frac{\sum_{i=1}^m f_i * y_i}{n}$$

Donde: $y_i = [\text{Límite superior intervalo } (i) - \text{Límite inferior intervalo } (i)] / 2$

EJEMPLO 10. Para ejemplificar se tomarán los 50 datos del ejemplo 3, del tiempo de atención en la Caja de un banco, que aparecen en la siguiente tabla:

OBS No.	TIEMPO (minutos)								
1	18	11	13	21	24	31	15	41	16
2	12	12	16	22	14	32	21	42	10
3	15	13	15	23	19	33	18	43	21
4	19	14	21	24	21	34	22	44	20
5	21	15	18	25	15	35	15	45	16
6	24	16	23	26	24	36	17	46	19
7	13	17	20	27	16	37	23	47	11
8	19	18	14	28	18	38	20	48	16
9	16	19	15	29	22	39	18	49	18
10	26	20	18	30	19	40	15	50	21

Utilizando la primera fórmula tiene que sumarse los 50 datos y dividirse entre 50, esto es como sigue:

$$(18 + 12 + 15 + \dots + 16 + 18 + 21) / 50 = 18 \text{ minutos}$$

Utilizando la forma agrupada en intervalos se utiliza la Tabla de Distribución de Frecuencias ya calculada:

Se destaca que existen varias formas de calcular la amplitud del intervalo que se va a utilizar. Una de estas variantes es tomar el mayor valor del conjunto de datos restarle el valor menor del conjunto de datos y el valor resultante se divide por 20 y por cinco, con esos valores se forma un intervalo, el obtenido en la división del primero punto y coma el valor obtenido de la división de segundo. De ese intervalo se selecciona un número, preferiblemente entero e impar, siempre que sea posible. Los dos números seleccionados no es “magia” obedece a que el investigador no quiere más de 20 clases ni menos de 5, porque en el primer caso sigue una gran cantidad de información y en el segundo, por lo contrario, es decir, por la pérdida de la información.

Intervalo	Frecuencia Absoluta (F _i)	Punto medio del intervalo (y _i)	F _i *y _i
[10 – 13]	5	(10+13)/2= 11.5	5*11.5= 57.5
(13 – 16]	15	14.5	217.5
(16 – 19]	13	17.5	227.5
(19 – 22]	11	20.5	225.5
(22 – 25]	5	23.5	117.5
(25 – 28]	1	26.5	26.5
	50		872

$$\bar{x} = 872/50 = 17.5 \text{ minutos}$$

La primera fórmula resulta exacta y la segunda es aproximada, debido a que en la medida que se agrupa, se pierde en exactitud, en la información.

Lo importante es comprender, que, a partir de un número, que está o debe estar en el centro del conjunto de los datos, se puede representar a todo el conjunto. Este número deja de cumplir esta condición, si existen valores extremos, en el conjunto de datos, lo cual falsea los resultados.

Con el desarrollo de la informática el cálculo del promedio se hace directamente a partir de los datos originales.

Utilizando el EXCEL se utiliza la Función Estadística PROMEDIO. Para ello se copia todos los datos en una Hoja EXCEL y se marca cualquier casilla y en las funciones estadísticas de EXCEL se selecciona Promedio y aparece una ventana, en la casilla de datos (Número 1) se coloca el rango donde están los datos que se desea promediar y aparece en la ventana el valor promedio (18), tal como se muestra en la siguiente figura 54.

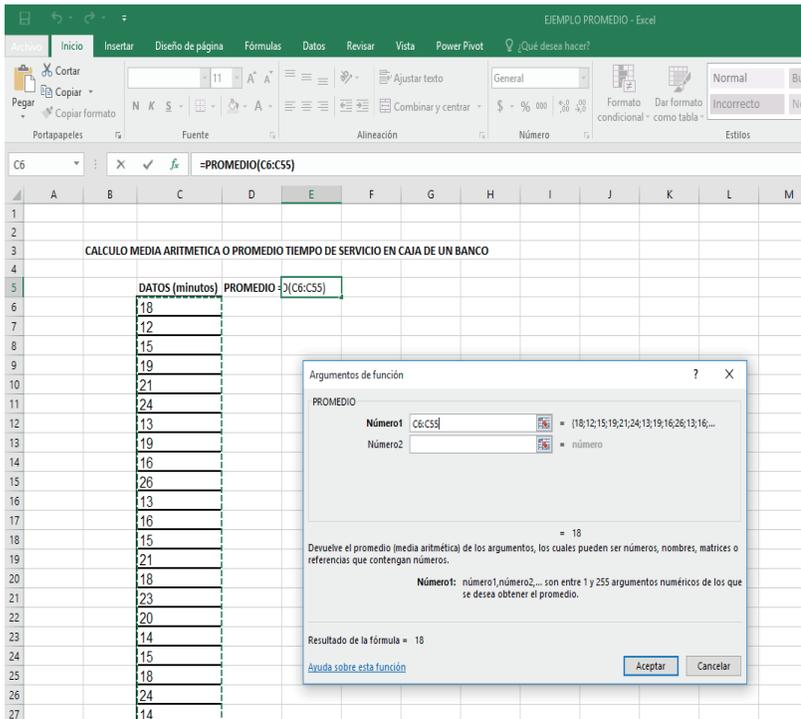


Figura 54. Cálculo de la Media Aritmética o Promedio utilizando EXCEL.

Cuando se marca Aceptar aparecerá el valor Promedio en la celda seleccionada de la Hoja EXCEL

La Media Aritmética o Promedio, como estadígrafo, proporciona un valor de uso muy común en cualquier tipo de estudios estadísticos, pero resulta conveniente precisar algunas propiedades o características de la misma que deben ser tenidos en consideración:

1. La suma de las *diferencias de la variable con respecto a la media* es nula, es decir,

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Resulta su uso conveniente cuando los datos se concentran simétricamente con respecto a ese valor.

2. Es el centro de gravedad de los datos.
3. Es el punto equidistante a los extremos.

Utilizando el SPSS se puede observar el siguiente resultado:

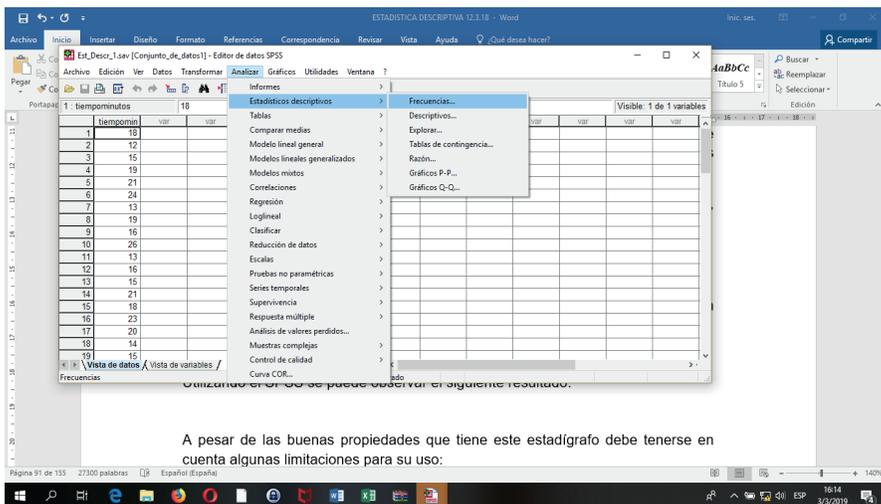


Figura 55. Panel analizar en el cálculo de la media aritmética o promedio utilizando SPSS.

Pasos: Utilizando la base de datos propia del ejemplo que se está tratando

1. Camino: Analizar/estadísticos descriptivos/frecuencia y aceptar

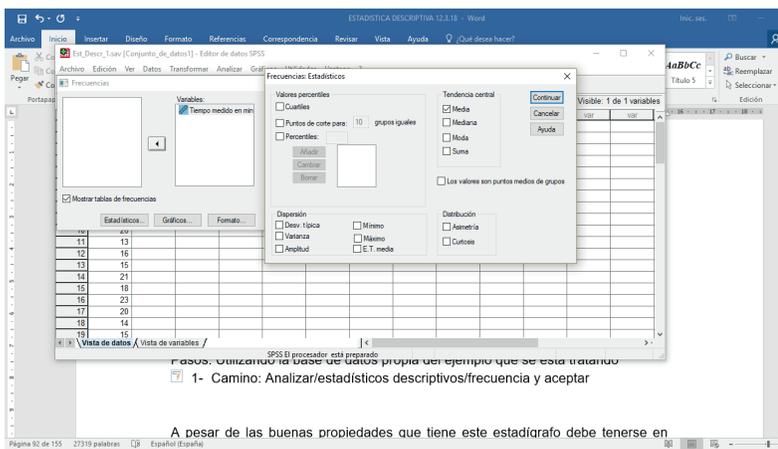


Figura 56. Selección del estadístico media aritmética o promedio utilizando SPSS.

2. Camino: seleccionar la variable indicada/estadígrafo/seleccionar/medidas de tendencia central/ media luego continuar/aceptar

Tiempo medido en minutos

N	Válidos	50
	Perdidos	0
Media		18.00

3. Se genera en resultados la tabla anterior que indica una media aritmética de 18, el mismo resultado del Excel, donde la interpretación es que de los pacientes que han llegado al banco al servicio de cajas, el tiempo de atención en el servicio es de 18 minutos.

A pesar de las buenas propiedades que tiene este estadígrafo debe tenerse en cuenta algunas limitaciones (González, 2013) para su uso:

1. La Media Aritmética o Promedio, resulta sensible a valores extremos, pues todas las observaciones intervienen en el cálculo de la misma. La aparición de observaciones extrema, hará que la media se desplace hacia esa dirección. En consecuencia; no es recomendable usar la media como medida de tendencia central en aquellas distribuciones, que, por su forma, resultan asimétricas.
2. Si consideramos una variable discreta, el valor de la media puede no pertenecer realmente al conjunto de los números enteros, ya que puede resultar un valor decimal.

2.1.2. Mediana

La Mediana es otro estadígrafo de tendencia central que tiene uso en el campo de la Estadística.

Para encontrar la Mediana de una muestra, que se denotará por M_{ed} , los valores de la muestra se ordenan de forma ascendente o descendente (de menor a mayor o viceversa). La Mediana es el valor de la observación cuya posición en la lista ordenada tiene igual número de observaciones por arriba y por debajo, es decir, es aquella que deja por debajo y por encima el 50% de los datos.

Si el tamaño de la muestra es $2n+1$, es decir, un valor impar, esta posición puede determinarse por $(n+1)/2$ y la Mediana es:

Med = Valor en la Posición $(n+1)/2$

Por ejemplo, si hay 25 observaciones en la muestra ordenada, la mediana es:

Med = Valor en la Posición $(25+1)/2 =$ Valor que ocupa la Posición 13

En el caso que el número de observaciones sea par, $2n$, entonces se toma como Mediana el valor promedio de las observaciones que están en la posición central. Estas dos posiciones será el par $n/2$ y $(n/2)+1$

Med = $[\text{Valor en la posición } n/2 + \text{Valor en la posición } [(n/2)+1]]/2$

Por ejemplo, si hay 32 observaciones, la Mediana es:

Med = $[\text{Valor en la posición } 32/2 + \text{Valor en la posición } [(32/2) + 1]]/2$

$[\text{Valor en la posición } n/2 + \text{Valor en la posición } [(n/2)+1]]/2$

Med = $(\text{Valor en la posición } 16 + \text{Valor en la posición } 17) / 2$

EJEMPLO 11. Se toman las primeras 20 observaciones del ejemplo del tiempo de atención en la Caja de un banco.

18	12	15	19	21	24	13	19	16	26
13	16	15	21	18	23	20	14	15	18

Estas observaciones ordenadas en forma ascendente quedarán como:

12	13	13	14	15	15	15	16	16	18
18	18	19	19	20	21	21	23	24	26

Como el número de observaciones es par, se toman los dos valores centrales que están en la posición 10 y 11 pues ambos dejan 9 observaciones por arriba y por debajo, respectivamente, entonces:

$M_{ed} = [18 (\text{posición } 10) + 18 (\text{posición } 11)] / 2 = (18 + 18)/2 = 36/2 = 18 \text{ min}$

Desarrollo del ejemplo a través del SPSS.

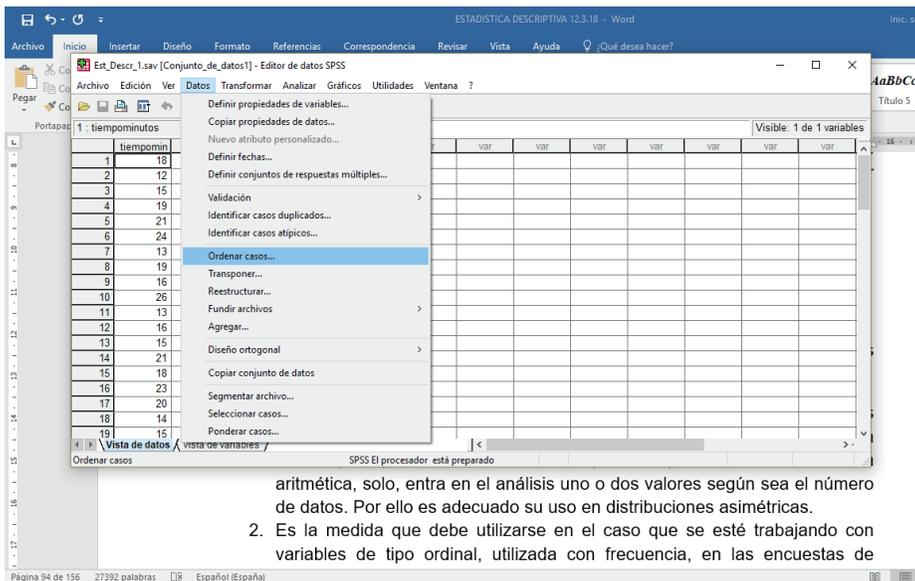


Figura 57. Ordenar casos en una variable utilizando SPSS.

Pasos:

1. Datos/ordenar casos/ retorno

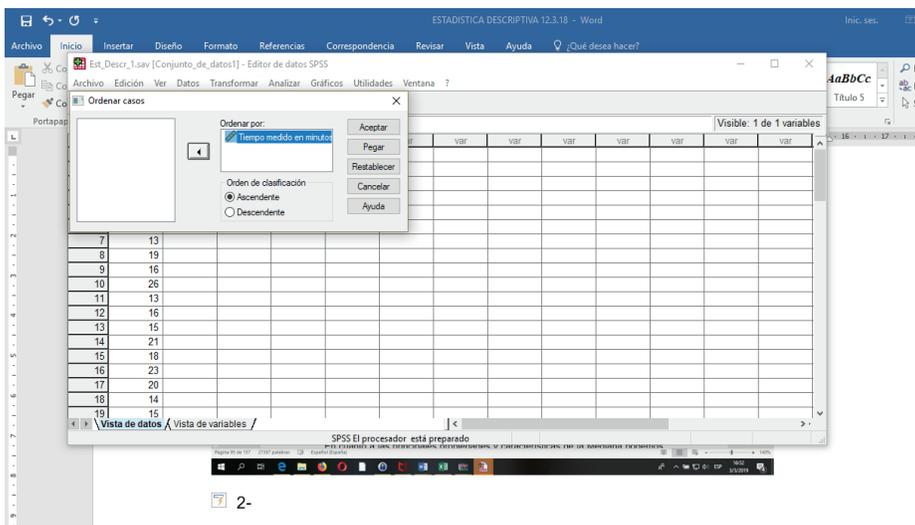


Figura 58. Selección de la variable que se va a ordenar ascendente utilizando SPSS.

2. Selección de la variable que se está trabajando/ascendente/ aceptar

Una vez ordenados en forma ascendente, en este caso, se procede a seleccionar los 20 primeros casos, el cual para hacerlo se puede utilizar diferentes vías.

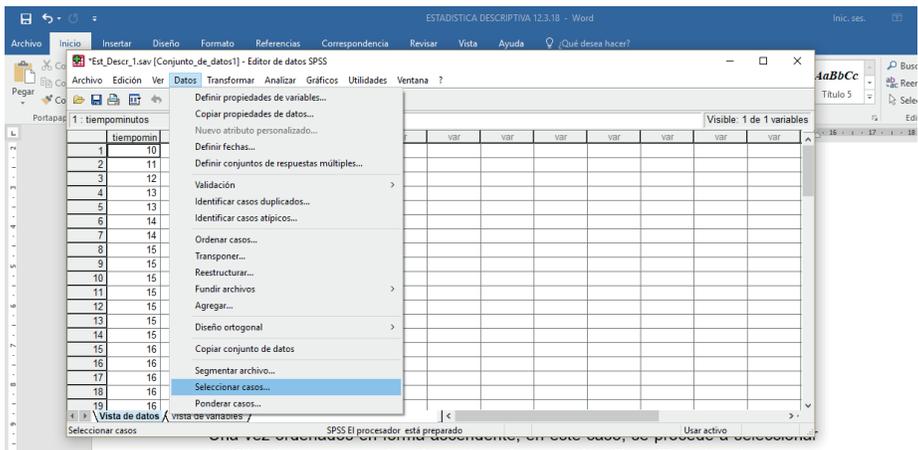
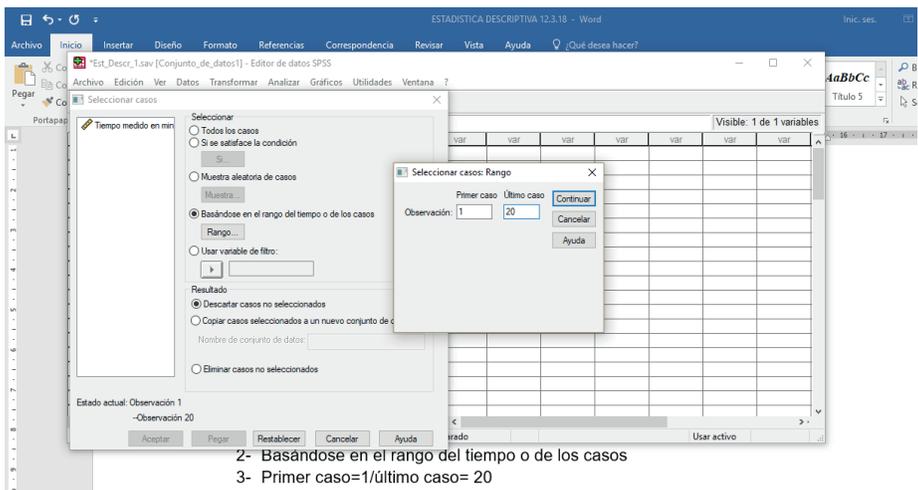


Figura 59. Panel de datos para seleccionar casos utilizando SPSS.

Pasos:

1. Datos/seleccionar casos



- 2- Basándose en el rango del tiempo o de los casos
- 3- Primer caso=1/último caso= 20

Figura 60. Seleccionar "Basándose en el rango del tiempo o de los casos" utilizando SPSS.

1. Basándose en el rango del tiempo o de los casos
2. Primer caso=1/último caso= 20/ continuar/aceptar

Los pasos son los mismos que se utilizan para calcular la media aritmética cambia la selección del estadístico como se muestra en la siguiente ventana

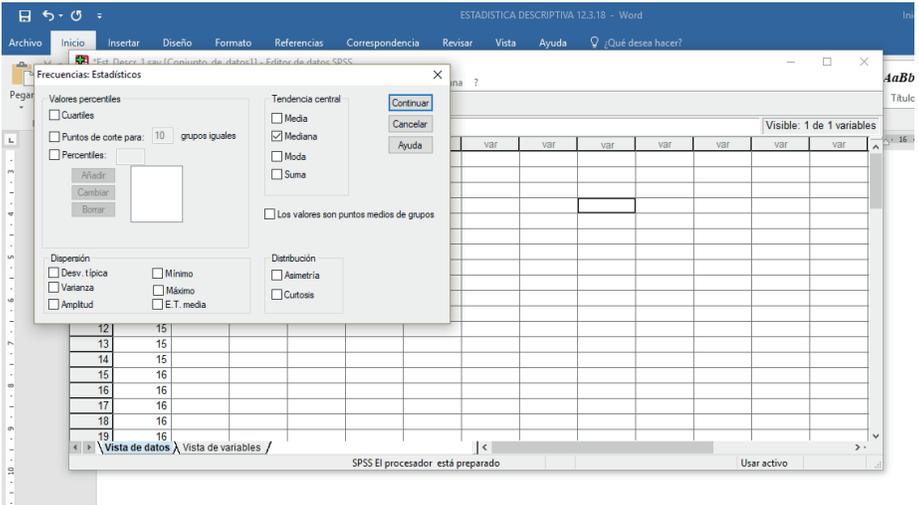


Figura 61. Selección del estadístico de la mediana utilizando SPSS.

Resultados

Tiempo medido en minutos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	10	1	5.0	5.0	5.0
	11	1	5.0	5.0	10.0
	12	1	5.0	5.0	15.0
	13	2	10.0	10.0	25.0
	14	2	10.0	10.0	35.0
	15	7	35.0	35.0	70.0
	16	6	30.0	30.0	100.0
Total		20	100.0	100.0	

Tiempo medido en minutos

N	Válidos	20
	Perdidos	0
Mediana		15.00

Cómo se observa el resultado es el mismo que utilizando el Excel, se ordenaron los 20 casos y como el conjunto era par se escogió los casos 10 (15) y 11(15), se sumaron y dividieron por dos $[(15+15)/2]$.

En cuanto a las principales propiedades y características de la Mediana podemos significar las siguientes:

1. Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino, del orden de las mismas, es decir, a diferencia de la media aritmética, solo, entra en el análisis uno o dos valores según sea el número de datos. Por ello es adecuado su uso en distribuciones asimétricas.
2. Es la medida que debe utilizarse en el caso que se esté trabajando con variables de tipo ordinal, utilizada con frecuencia, en las encuestas de estudios de problemas de administración y social.
3. Es utilizable en algunos métodos de Inferencia Estadística, pero no con la amplitud que se utiliza la Media Aritmética o promedio, pues la estadística que subyace sobre ella, es no paramétrica, que es menos potente que la paramétrica.

EJEMPLO 12. La Mediana puede ser fácilmente calculada utilizando el EXCEL. Para ello se utilizará la misma hoja EXCEL usada para el cálculo del Promedio, en las Funciones Estadísticas se selecciona la función MEDIANA y aparece una ventana similar a la del promedio; en la casilla de los datos, se coloca el rango de valores para el cálculo de la Mediana, tal como se muestra en la siguiente figura.

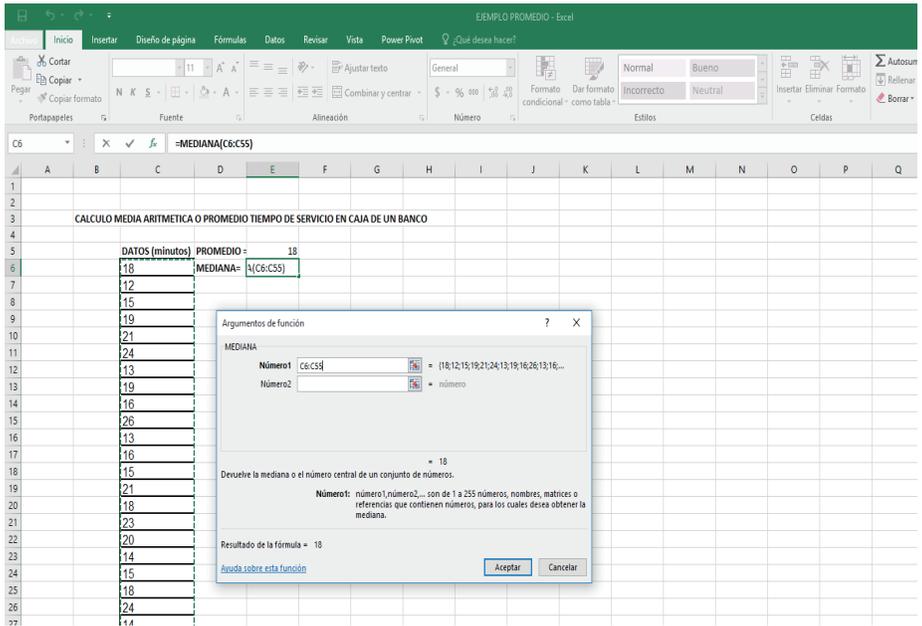


Figura 62. Ventana para el cálculo de la mediana, usando el Excel.

Directamente en la ventana aparece el valor de la Mediana (18) y cuando se marca Aceptar, este valor aparecerá en la celda seleccionada de la Hoja EXCEL.

2.1.3. Trimedia

La trimedia es una medida robusta de tendencia central que tiene sus ventajas sobre la media y la mediana, es decir, ella representa mejor el conjunto de los datos pues trabaja con todos los valores menos con los extremos.

$$T_x^- = \frac{Q_1 + 2Q_2 + Q_3}{4}$$

Donde:

Q_1 = Cuartil 1= 25% de los datos

Q_2 = Cuartil 2= Media aritmética

Q_3 = Cuartil 3= 75% de los datos

En el ejemplo 3, que se ha estado tratando en el texto, se tomaran los 50 datos, del tiempo de atención en la Caja de un banco, se desea calcular la trimedia de los mismo, con ayuda del SPSS

Los pasos son idénticos a los utilizados en el cálculo de la media y la mediana

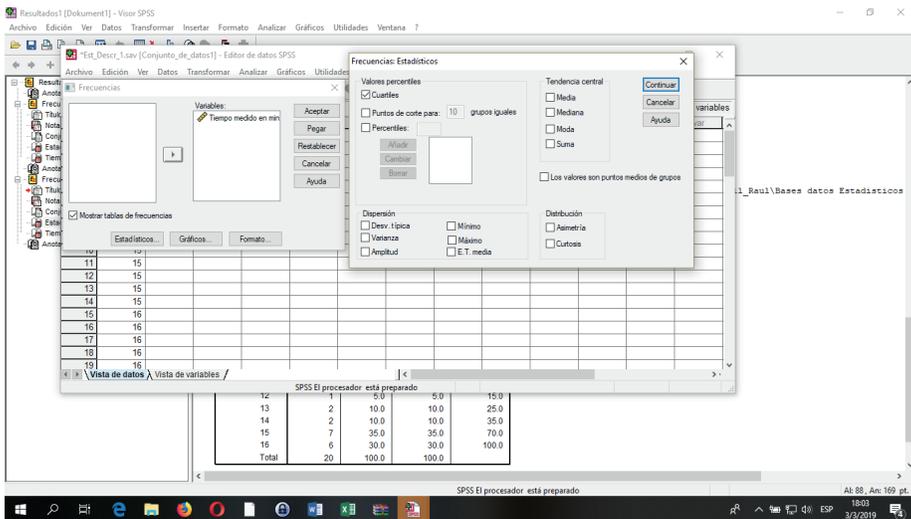


Figura 63. Ventana para seleccionar los cuartiles utilizando SPSS.

Pasos:

Selección de la variable/valores percentiles/cuartiles/continuar/aceptar

Estadísticos

Tiempo medido en minutos

N	Válidos	50
	Perdidos	0
Percentiles	25	15.00
	50	18.00
	75	21.00

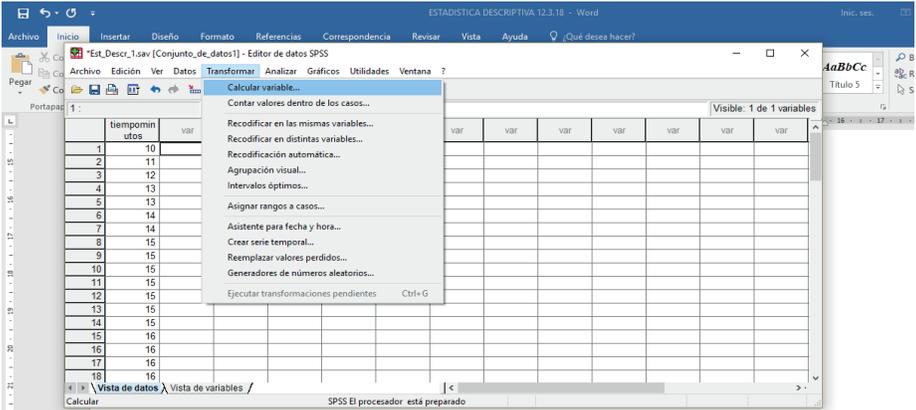
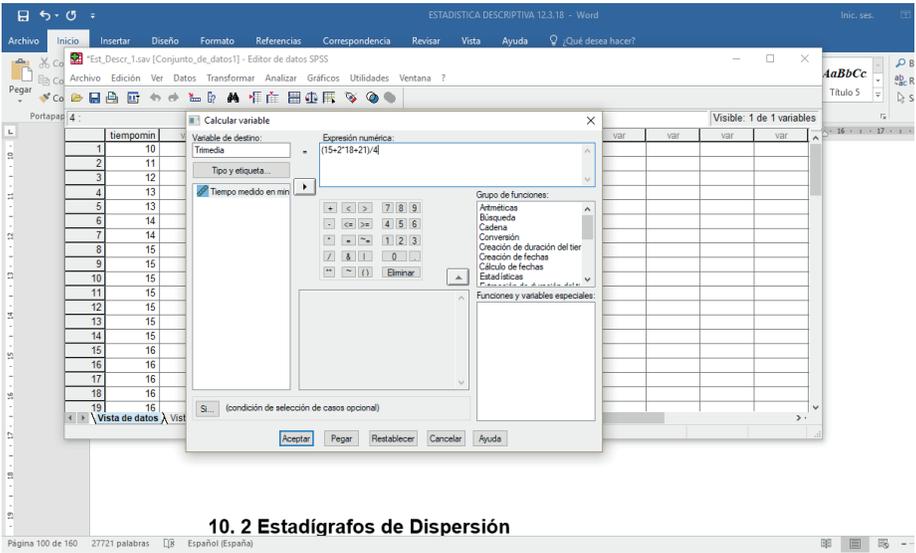


Figura 64. Ventana para calcular una variable utilizando SPSS.

Pasos: Transformar/ calcular variables



10. 2 Estadígrafos de Dispersión

Figura 65. Ventana para la operación calcular una variable utilizando SPSS.

Pasos:

1. colocar un nombre a la variable, en este ejemplo "Trimedia"
2. expresión numérica/ sustituir el valor de los cuartiles según fórmula/
Aceptar

Resultado:

$$T_x^- = 18$$

Cómo se puede apreciar, en este ejemplo particular, la media coincide con la trimedia y con la mediana, aunque, esta última, no se haya calculado con el total de los datos, en este texto. En el próximo libro se verá la importancia de esta coincidencia, es decir, de la coincidencia de los estadísticos de tendencia central.

2.1.4. Moda

Es el valor de la variable que tenga mayor Frecuencia Absoluta. En el caso de variables numéricas, si se organiza por intervalos, se identifica como Intervalo Modal, que será el Intervalo con mayor Frecuencia Absoluta.

EJEMPLO 13. En el Ejemplo 1 sobre el tipo de turismo en una Agencia de Viajes la tabla de Frecuencia Absoluta fue:

CLASE	CANTIDAD
Sol y Playa (SP)	33
Ecoturismo (ET)	25
Turismo de aventura (TA)	12
TOTAL	70

Por tanto, la Moda de este problema de variable categórica es la variable Sol y Playa que es la de mayor Frecuencia Absoluta (33).

En el caso del ejemplo del tiempo de atención en la Caja de un Banco, que es una variable numérica, cuya tabla de Frecuencia Absoluta se mostró anteriormente (Tabla 4) el Intervalo Modal es de 13 a 15 minutos, ya que tiene la mayor Frecuencia Absoluta (15).

De la moda podemos destacar las siguientes propiedades:

1. Es muy fácil de calcular.
2. Puede no ser única dentro de una misma distribución o incluso no existir.
3. Tiene fundamentalmente interés en la Estadística Descriptiva y para variables categóricas.

Para poder utilizar el EXCEL para el cálculo de la Moda, en el caso de las variables categóricas o cualitativas, se requiere que ellas aparezcan codificadas en forma numérica.

El cómo hacer el cálculo de este estadígrafo, con la utilización del Excel, se verá en los problemas resueltos sobre esta temática.

2.2. Estadígrafos de Dispersión

Son aquellos estadígrafos que indican la mayor o menor concentración de los datos con respecto a las medidas de tendencia central. En este caso se estudiarán los siguientes: **Amplitud o Rango, Varianza, Desviación Típica o Estándar y Coeficiente de Variación.**

2.2.1. Amplitud o Rango

Una medida razonable de la variabilidad podría ser la **Amplitud o Rango**, que se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto. Es fácil de calcular y sus unidades son las mismas que las de la variable.

RANGO = DATO CON MAYOR VALOR – DATO CON MENOR VALOR

Cuando se estudió la determinación de los intervalos para agrupar datos numéricos, el primer paso era calcular el Rango. Mientras mayor sea el Rango o Amplitud de los datos, mayor es su variabilidad, o sea están más dispersos.

Este estadígrafo posee varios inconvenientes:

- No utiliza todas las observaciones (solo dos de ellas);
- Se puede ver muy afectada por alguna observación extrema;
- El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso, nunca disminuye.

2.2.2. Varianza Muestral

La Varianza Muestral, S^2 , se define como la media de las diferencias cuadráticas de n valores (x_i) con respecto a su Media Aritmética o Promedio. Esta puede ser calculada para una muestra por:

$$S^2 = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Una fórmula equivalente para el cálculo de la varianza es:

$$s^2 = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

La varianza no tiene la misma magnitud que las observaciones, por ejemplo, si las observaciones se miden en metros, la varianza estaría en metros cuadrados (m²).

Si se quiere calcular la Varianza para la población completa (Varianza Poblacional), se harían algunas variaciones a la fórmula anterior, quedando como:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Y su fórmula equivalente será:

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n}$$

2.2.3. Desviación Típica o Estándar Muestral

Esta medida de dispersión se deriva directamente de la Varianza y es de las más utilizadas en la Estadística. La Desviación Típica o Estándar, que se denota por S, se define como:

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

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Esto permite que la medida de la dispersión o variabilidad de los datos esté en la misma dimensión que tienen estos y por ende se facilitan los análisis. Si, por ejemplo, se estuviera analizando el comportamiento de un equipo de llenado de envases de café molido de 250 gramos y se desea conocer cuánto se está desviando de este peso, se calcularía la Desviación Típica tomando como promedio el valor de 250 y está daría los gramos que se desvía por debajo y por arriba de los 250 gramos.

EJEMPLO 14. Para ejemplificar el uso de las fórmulas, se tomarán las 10 primeras observaciones del ejemplo del tiempo de atención en la Caja de un banco. Los datos son:

18	12	15	19	21	24	13	19	16	26
----	----	----	----	----	----	----	----	----	----

Para calcular la varianza se utiliza la siguiente fórmula:

$$S^2 = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}$$

$$\sum_{i=1}^n x_i^2 = 18^2 + 12^2 + 15^2 + 19^2 + \dots + 16^2 + 26^2 = 3\,533$$

$$\sum_{i=1}^n x_i = 18 + 12 + 15 + 19 + 21 + 24 + 13 + 19 + 16 + 26 = 183$$

$$S^2 = [10(3\,533) - 183^2] / 90 = 20.45 \text{ min}^2$$

Entonces la Desviación Típica o Estándar será:

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

$$s = \sqrt{20.45 \text{ min}^2}$$

$$S = 4.52 \text{ min}$$

Tanto la Varianza Muestral como la Desviación Típica tienen una serie de propiedades y características que resultan fundamentales en materia de Inferencia Estadística. Entre ellas cabe destacar

1. Ambas son sensibles a la variación de cada una de los datos, es decir, si un dato cambia, cambia con ello la varianza. La razón es que si miramos su definición, la varianza es función de cada uno de los datos.
2. A medida que el tamaño de la muestra se incrementa, ambas disminuyen, porque son inversamente proporcionales al valor de n, o sea cuando n crece, la dispersión o variabilidad con respecto al Promedio disminuye.

2.2.4. Coeficiente de Variación

El Coeficiente de Variación es un estadígrafo que permite medir la relación existente entre la variabilidad, medida por la Desviación Típica y el valor central de una variable, medida por su promedio y se expresa por:

$$CV = \frac{S}{\bar{x}} * 100$$

La interpretación es que a medida que el Coeficiente de Variación es más pequeño, los valores de la variable están más concentrados alrededor del valor promedio. Si se estuviera estudiando una variable que tiene siempre el mismo valor, la Desviación Estándar sería cero y por tanto cero el Coeficiente de Variación. En algunas situaciones prácticas, por ejemplo, el estudio del comportamiento de la demanda de un cierto producto, si el Coeficiente de Variación es pequeño ($CV \leq 10$) se puede asumir que esta demanda es prácticamente constante. De forma similar se utiliza para hacer comparaciones de la variabilidad entre variables que tengan distintas dimensiones. Por ejemplo, si se quisiera comparar la variabilidad del peso de la soya en una hectárea de cultivo, que se midiera en kilogramos y el rendimiento de esta en aceite que se mediría en litros.

EJEMPLO 15. Tomando los datos de las 10 primeras observaciones del tiempo de atención en la Caja de un banco, visto anteriormente, el Coeficiente de Variación se calcula como:

$$CV = \frac{S}{\bar{x}} * 100$$

$$\bar{x} = \frac{183}{10} = 18.3 \text{ minutos}$$

$$S = 4.52 \text{ minutos}$$

$$CV = (4.52/18.3)100 = 24.69\%$$

Por lo que puede concluirse que la dispersión de esta variable con respecto al promedio no es grande.

Los estadígrafos estudiados anteriormente pueden ser fácilmente calculados utilizando el EXCEL.

EJEMPLO 16. Para ejemplificar se tomará la Tabla de datos completa del ejemplo del tiempo de atención en la Caja de un banco, la cual es:

EJEMPLO TIEMPO DE ATENCION CAJA DE UN BANCO

DATOS

18	13	24	15	16
----	----	----	----	----

12	16	14	21	10
15	15	19	18	21
19	21	21	22	20
21	18	15	15	16
24	23	24	17	19
13	20	16	23	11
19	14	18	20	16
16	15	22	18	18
26	18	19	15	21

La matriz anterior se copia en una hoja EXCEL y se prepara una pequeña tabla para calcular la varianza y la desviación típica o estándar. La hoja queda según la figura 66.

The screenshot shows an Excel spreadsheet with the following content:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1																			
2																			
3																			
4																			
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			

Labels in the spreadsheet:

- Row 6: DATOS
- Row 7: 18 13 24 15 16
- Row 8: 12 16 14 21 10
- Row 9: 15 15 19 18 21
- Row 10: 19 21 21 22 20
- Row 11: 21 18 15 15 16
- Row 12: 24 23 24 17 19
- Row 13: 13 20 16 23 11
- Row 14: 19 14 18 20 16
- Row 15: 16 15 22 18 18
- Row 16: 26 18 19 15 21

Calculation labels:

- Row 7: CALCULO DE LOS ESTADIGRAFOS DE DISPERSION O VARIABILIDAD
- Row 8: VARIANZA
- Row 9: DESV TIPICA

Figura 66. Hoja de Excel para el cálculo de la varianza y la desviación típica.

Para el cálculo de cada uno de estos estadígrafos existe una función en EXCEL. Por ejemplo, si se desea calcular la varianza para los datos en la tabla, entonces se marca la celda a la derecha del título de VARIANZA en la hoja EXCEL, se selecciona la función de la Varianza Muestral (VAR.S) y aparece la ventana para su cálculo. En la casilla para Numero 1 se coloca la matriz de datos del problema y ya aparece en la misma ventana el valor de la varianza (13.14). Esto se muestra en la siguiente figura 67.

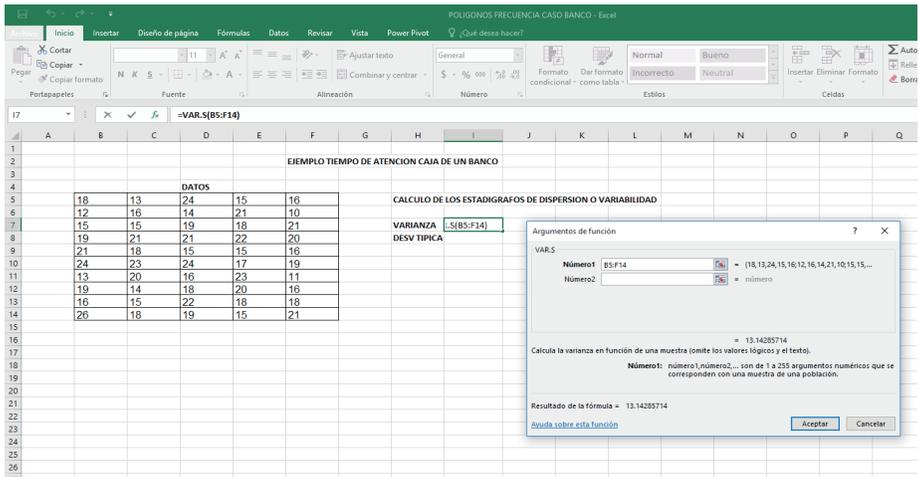


Figura 67. Ventana de Excel para cálculo de la varianza.

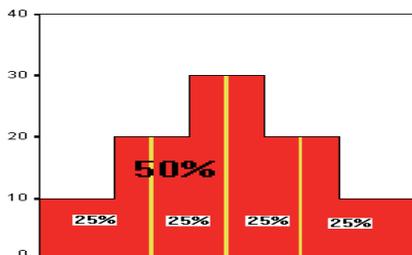
Cuando se da aceptar aparecerá este valor en la celda seleccionada.

De forma similar puede calcularse la Desviación Típica, por medio de la Función DESVEST.M (Desviación Estándar Muestral).

2.3. Estadígrafos de Posición

Los estadígrafos de posición son aquellos que dividen un conjunto ordenado de datos en grupos con una determinada cantidad de individuos. De ellos se estudiará los cuartiles y percentiles.

Un caso importante en este tipo de estadígrafos de posición son los denominados **cuartiles**, que son aquellos que dividen la distribución ordenada ascendente de los datos en 4 partes iguales correspondientes, por tanto, cada uno de ellos al 25% de los elementos de la población. Gráficamente estos se representan como:



Los Cuartiles se denotan por:

Q_1 : El valor de la variable por debajo del cual hay un 25% de las observaciones o datos.

Q_2 : El valor de la variable por debajo del cual hay un 50% de las observaciones o datos.

Q_3 : El valor de la variable por debajo del cual hay un 75% de las observaciones o datos.

Note que el 2do Cuartil (Q_2) coincide con el valor de la Mediana estudiada anteriormente.

Para la determinación de la posición en la serie ordenada de n valores, que corresponde a cada Cuartil, pueden utilizarse las siguientes fórmulas:

$Q_1 =$ Valor que aparece en la posición $(n+1)/4$

$Q_2 =$ Valor que aparece en la posición $(n+1)/2$

$Q_3 =$ Valor que aparece en la posición $3(n+1)/4$

Debe tenerse en cuenta las siguientes reglas:

- Si el valor de la posición es un entero, se toma esa posición para seleccionar el Cuartil. Ejemplo si hay 15 observaciones, el Q_1 estaría en la posición $(15+1)/4$, o sea en la posición 4, el Q_2 en la 8va posición y Q_3 en la posición 12.
- Si el valor de la posición es el punto medio entre dos enteros, se toma el promedio del valor de la variable que está en la posición de los dos enteros. Ejemplo, si hay 17 observaciones, la posición para Q_1 sería $(17+1)/4 = 4.5$,

Entonces el valor del 1er Cuartil, será el promedio de la variable que está en la posición 4 y la 5; el Q_2 estaría en la posición $(17+1)/2$, o sea es el valor de la variable que aparece en la posición 9 y Q_3 estaría en la posición $3(17+1)/4=13.5$ y se promedia los valores de la variable que está en la posición 13 y 14.

- Si el valor de la posición da un valor decimal que no es el punto medio entre dos enteros, entonces el valor del Cuartil se calcula por la fórmula:

$$Q_j = X_i + d*(X_{i+1} - X_i)$$

- Por ejemplo si hay 14 observaciones, la posición de Q_1 se determinaría como $(14+1)/4=3.75$, entonces Q_1 es el valor:

$$Q_1 = X_3 + 0.75(X_4 - X_3)$$

- Q_2 está en la posición $15/2=7.5$ y es el promedio del valor de la variable que aparece en la posición 7 y la 8.
- Q_3 está en la posición $3(15)/4=11.25$ y se calcula por:

$$Q_3 = X_{11} + 0.25(X_{12} - X_{11})$$

Los Cuartiles son las medidas de posición que más se utilizan en la Estadística Descriptiva y a partir de ellos se construyen gráficos de mucha utilidad para poder comparar varias muestras de una misma población, como se estudiará posteriormente.

El otro estadígrafo de posición que se estudiará son los **percentiles**, que son los valores que dividen los datos de una serie ordenada en forma ascendente, en cien partes iguales o en los que cada uno de ellos concentra el 1% de los resultados. Por tanto, el Percentil 90, será el valor de la variable que tiene un 90% de los datos por debajo.

Partiendo de estas consideraciones tendríamos que:

Q_1 = Primer cuartil = Percentil 25

Q_2 = Segundo cuartil = Percentil 50

Q_3 = Tercer cuartil = Percentil 75

Los Percentiles C, se corresponde con el valor de la variable en una serie ordenada de forma ascendente que esté en la posición $(C/100)*n$

EJEMPLO 17. Para ejemplificar se tomarán las primeras 20 observaciones del tiempo de atención en la Caja de un banco, cuya serie ordenada vista anteriormente es:

12	13	13	14	15	15	15	16	16	18
18	18	19	19	20	21	21	23	24	26

Para calcular los Cuartiles:

Q_1 es el valor de la variable en la posición $(n+1)/4= 21/4=5.25$. Al ser un valor decimal, este Cuartil se calcula por:

$$Q_1 = 15 + 0.25(15 - 15) = 15$$

Q_2 está en la posición $21/2=10.5$. Como este valor es el punto medio entre 10 y 11, el segundo Cuartil será el promedio del valor de la variable en estas posiciones, por tanto:

$$Q_2 = (18+18)/2 = 18 \text{ minutos}$$

Q_3 está en la posición $3*21/4 = 15.75$ y se le calcula por:

$$Q_3 = 20 + 0.75(21 - 20) = 20.75$$

El Percentil 90 para este ejemplo está en la posición ordenada $0.90*20 = 18$ y por tanto es 23 minutos.

De forma similar el Percentil 80 es el valor de la variable en la serie ordenada que está en la posición $0.8*20 = 16$ y es 21 minutos.

Utilizando el EXCEL, los Cuartiles se pueden calcular fácilmente. Para ello se coloca en una Hoja EXCEL los datos. Se incluye en celdas el nombre de los estadígrafos de posición que se desea calcular, se selecciona la celda a la derecha del estadígrafo y se busca en Funciones Estadística del EXCEL la función CUARTIL.EXC y aparece la ventana para la entrada de datos. Esto se muestra en la Figura 68.

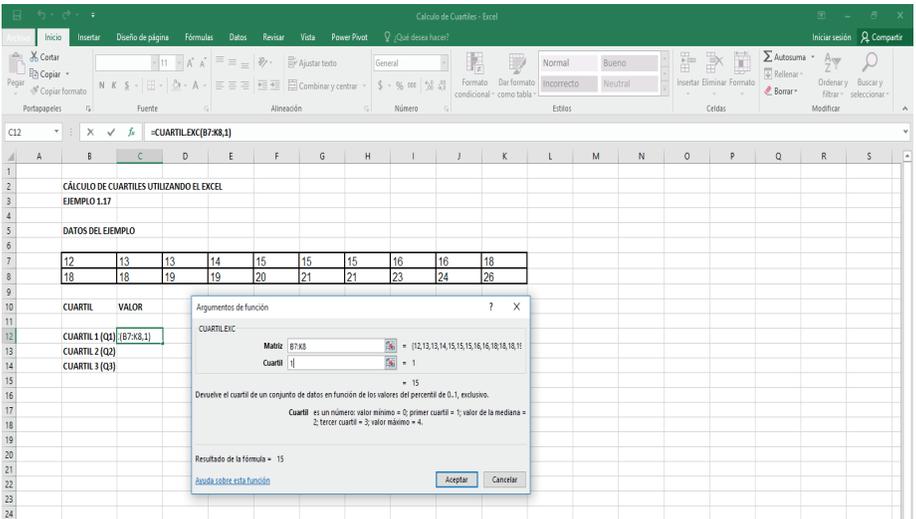


Figura 68. Ventana de Excel para el cálculo de los cuartiles.

En la casilla de Matriz, se coloca la matriz de los datos marcándola directamente y en la casilla de Cuartil, se coloca el valor del Cuartil que se desea calcular, en este caso el Cuartil 1 (Q_1). En la ventana, dado estos valores, aparece directamente el valor del Cuartil 1, que es de 15 minutos.

De forma similar se haría para el cálculo del resto de los Cuartiles y finalmente los resultados se muestran en la Figura 69.

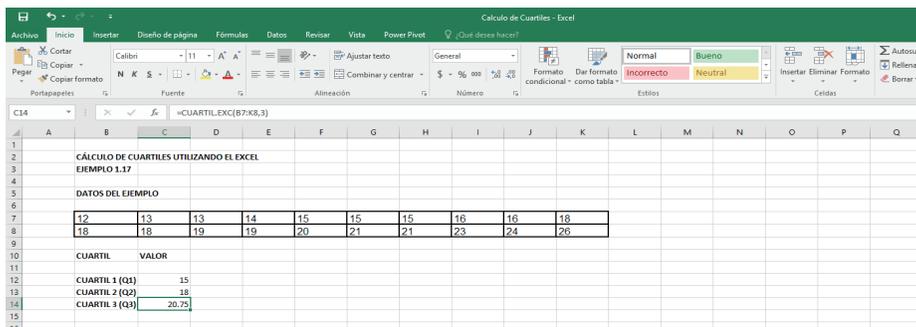


Figura 69. Hoja de Excel con los resultados del cálculo de los cuartiles.

2.3.1. Rango Intercuartil

Este estadígrafo permite medir la dispersión de los datos, pero basado en las medidas de posición.

Este estadígrafo se define como:

$$RIC = Q_3 - Q_1$$

Este rango mide la dispersión en el 50% de las observaciones que están en el centro de la distribución de datos, por tanto no es influenciado por los valores extremos que pueda tomar la variable. A mayor valor, mayor dispersión en los datos centrales. Más adelante se verá su utilización, cuando se desea comparar dos distribuciones de datos

2.3.2. Estadígrafos de forma o distribución

En cuanto a la forma en una distribución de los datos se estudiarán dos tipos de estadígrafos: estadígrafos de Asimetría y de Curtosis o Apuntalamiento:

El estadígrafo de Asimetría, nos permite saber si los datos se distribuyen de forma simétrica con respecto a un valor central, o si bien la gráfica que representa la distribución de frecuencias es de una forma diferente del lado derecho que del lado izquierdo.

Para saber si una distribución de frecuencias es simétrica, hay que precisar con respecto a qué. La mayoría de los estadígrafos que miden la asimetría utilizan la media como medida de tendencia central.

Dentro de los tipos de asimetría posible, se destacan las dos fundamentales:

Asimetría positiva: Si las frecuencias más altas se encuentran en el lado izquierdo de la mediana, mientras que en derecho hay frecuencias más pequeñas (cola).

Asimetría negativa: Cuando la cola está en el lado izquierdo.

Se puede representar los diferentes tipos de asimetría en las distribuciones de la forma siguiente:

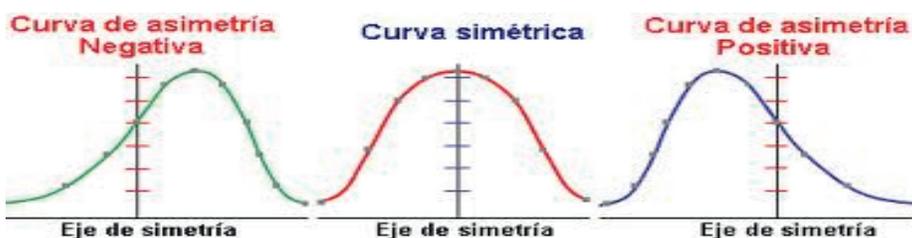


Figura 70. Tipos de asimetría según las distribuciones.

Los cálculos de indicadores de la simetría resultan relativamente complejos, por lo que solo se estudiará aquí su interpretación y su cálculo se realizará utilizando la informática, en este caso el EXCEL.

El Índice o Coeficiente de Simetría (G_s), se interpreta de la siguiente forma:

$G_s = 0$, implica una distribución simétrica; o sea, existe la misma concentración de valores a la derecha y a la izquierda de la media.

$G_s > 0$, resulta la expresión de una distribución asimétrica positiva; en la cual, por lo tanto, existe mayor concentración de valores a la derecha de la media que a su izquierda. En este caso el valor obtenido de este coeficiente está cercano a uno (1).

$G_s < 0$, como expresión de una distribución asimétrica negativa; existe mayor concentración de valores a la izquierda de la media que a su derecha. En este caso el valor obtenido de este coeficiente está cercano a menos uno (-1).

Apuntamiento o curtosis; a través de este tipo de estadígrafos podremos conocer si la curva es más o menos apuntada (larga y estrecha). Por lo que

el coeficiente o índice de curtosis analiza el grado de concentración que presentan los valores alrededor de la zona central de la distribución.

A partir de este tipo de indicadores se definen 3 tipos de distribuciones según su grado de curtosis:

Distribución Mesocúrtica: presenta un grado de concentración medio alrededor de los valores centrales de la variable (el mismo que presenta una distribución simétrica).

Distribución Leptocúrtica: presenta un elevado grado de concentración alrededor de los valores centrales de la variable.

Distribución Platicúrtica: presenta un reducido grado de concentración alrededor de los valores centrales de la variable.

Gráficamente se puede interpretar el índice de Curtosis de la manera siguiente:

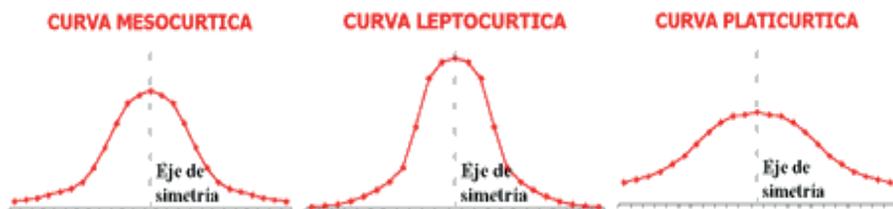


Figura 71. Tipos gráficos asociados a la curtosis.

Los indicadores calculados en estos casos podremos identificarlos como G_c ; en tal sentido sus valores deben ser interpretados de la forma siguiente:

$G_c = 0$; se interpreta como indicador de una distribución **mesocúrtica**.

$G_c > 0$; implica que estaríamos en presencia de una distribución **leptocúrtica**.

$G_c < 0$; indicativo de una distribución **platicúrtica**.

EJEMPLO 18. Para ejemplificar el cálculo de estos estadígrafos, se tomarán los datos del tiempo de atención en la Caja de un banco. Lo primero es copiar los datos en una Hoja EXCEL. Para calcular la Asimetría, se utilizará dentro de las funciones estadísticas, la identificada como COEFICIENTE.ASIMETRIA y se abrirá una ventana para introducir los datos, tal como se muestra en la siguiente figura 72.

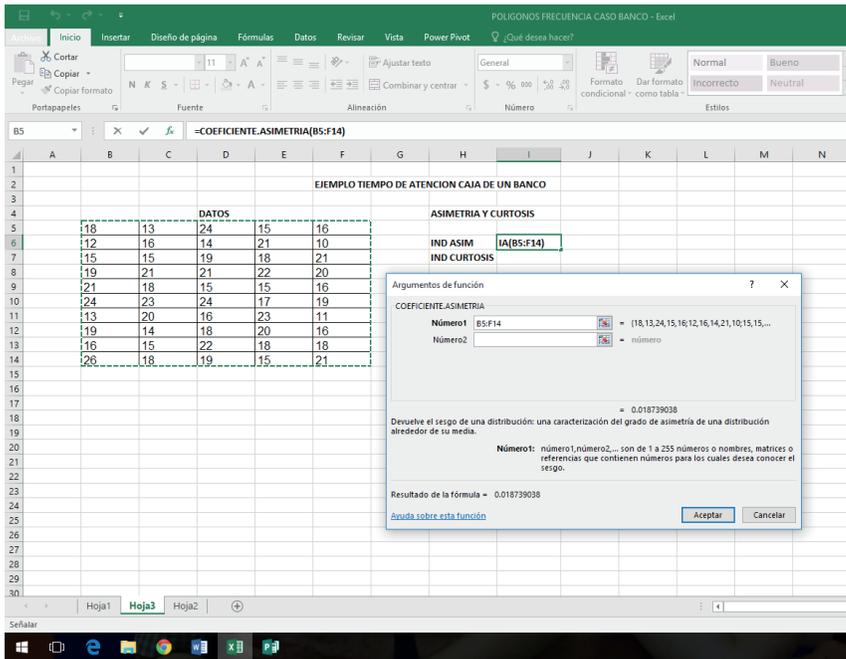


Figura 72. Ventana de Excel para el cálculo del coeficiente de asimetría.

En la casilla de Número 1 se introduce la matriz de los datos y de forma inmediata aparece el valor del Coeficiente o índice de Asimetría en este caso 0.0187

El Coeficiente o Índice de Asimetría es ligeramente positivo, pero muy cercano a cero, por lo que a los efectos prácticos la distribución de esta variable puede considerarse simétrica con respecto a la media.

En el caso de la curtosis el procedimiento sería similar y se utiliza la función estadística CURTOSIS, apareciendo una ventana similar a la anterior y colocando los datos en la casilla Número 1 se mostrará de inmediato en la misma ventana el valor del Coeficiente de Curtosis, tal como se muestra en la figura 73.

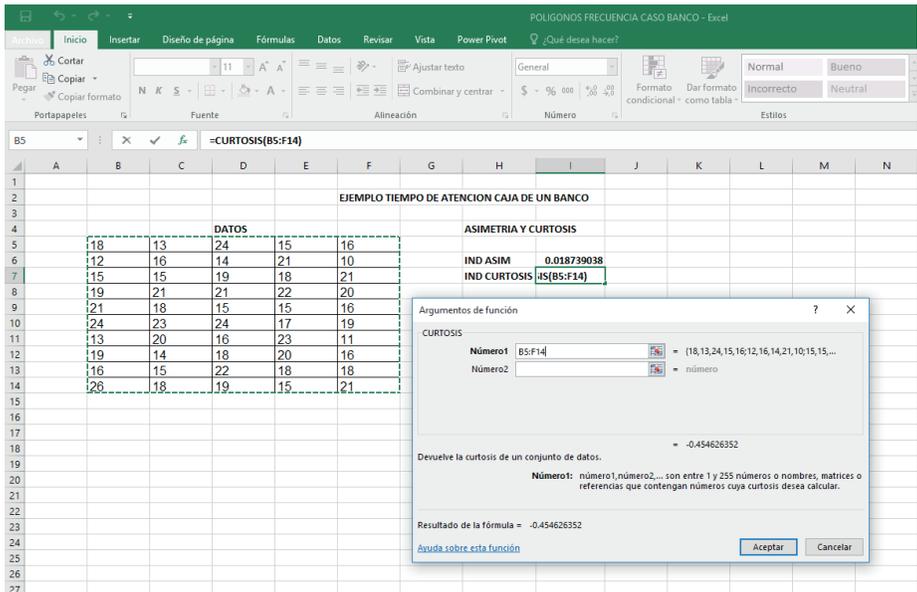


Figura 73. Ventana de Excel para el cálculo de la curtosis.

Como puede observarse el valor de ese coeficiente es de -0.4546, por lo que se considera que los datos tienen una distribución ligeramente platycúrtica.

Capítulo III. Análisis exploratorio de los datos usando los cuartiles. El diagrama de caja

3.1. El Diagrama de Caja en los análisis exploratorios

Como se estudió anteriormente, las representaciones gráficas pueden ayudar a visualizar de forma rápida los resultados de un experimento estadístico.

El Diagrama de Caja es un gráfico muy utilizado en los análisis exploratorios de los datos, que permite estudiar la tendencia central y la dispersión de los datos de un experimento con relación a su medida central. Su construcción resulta sencilla y se basa fundamentalmente en los Cuartiles anteriormente estudiados.

Este tiene forma rectangular y la longitud de este rectángulo es el Rango Intercuartil. Dentro del rectángulo aparece una línea que representa la mediana y se adicionan dos líneas, una a cada lado del rectángulo, denominadas popularmente bigotes. Éstos bigotes se construyen con una línea que parte del lado menor del rectángulo, hasta el menor valor de la muestra y del lado con mayor valor hasta el valor mayor de las observaciones. Los paquetes informáticos estadísticos permiten construir los Diagramas de Caja. En la figura 74 se representa un Diagrama de Caja:

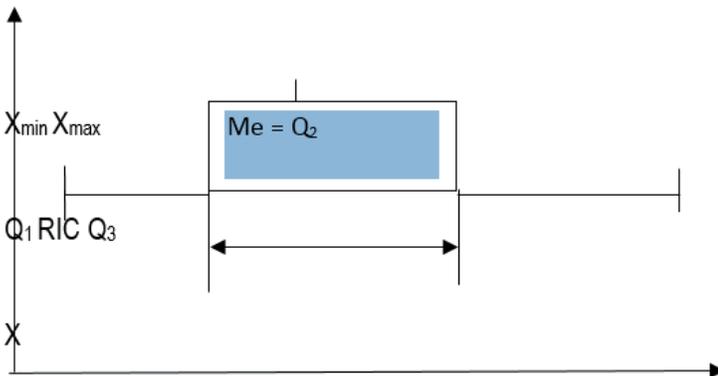


Figura 74. Distribución de los cuartiles y el rango intercuartil, en un diagrama de caja.

El Rango Intercuartil (RIC) se definió anteriormente como:

$$\text{RIC} = Q_3 - Q_1$$

A partir del Rango Intercuartil, pueden calcularse dos límites, que permiten identificar cuando pueden existir datos sospechosos o anormales (outliers) que deben ser investigados, pues afectan las medidas de tendencia central y de desviación y pueden deberse a situaciones anormales y es aconsejable eliminarlos de la muestra. Estos límites se calculan de la siguiente forma:

$$\text{LIS} = Q_1 - 1.5 \cdot \text{RIC} \text{ Límite inferior de sospecha}$$

$$\text{LSS} = Q_3 + 1.5 \cdot \text{RIC} \text{ Límite superior de sospecha}$$

$$\text{LIA} = Q_1 - 3.0 \cdot \text{RIC} \text{ Límite inferior de anormalidad}$$

$$\text{LSA} = Q_3 + 3.0 \cdot \text{RIC} \text{ Límite superior de anormalidad}$$

En caso que no haya datos sospechosos ni anormales los bigotes del Diagrama de Caja coincidirán con el valor mínimo y máximo de la muestra. Otros autores definen el punto de cierre del bigote como el Percentil 5% y el Percentil 95%, cuando hay muchos datos.

Este tipo de gráfico permite:

- Determinar la simetría de los datos de acuerdo a la posición de la Mediana en el rectángulo. Si la línea que representa la Mediana está cercana al centro del rectángulo los datos están distribuidos simétricamente. En caso de que aparezca hacia la derecha o izquierda habrá asimetría en la distribución de los datos.
- Determinar la dispersión o variabilidad de los datos. A medida que el rectángulo es más abierto, mayor es la dispersión de la información
- Poder comparar a priori, en dos o más muestras, sus diferencias en la tendencia central y en la dispersión.
- Poder visualizar el Rango en que se mueven los datos confiables, observando los extremos del bigote.
- Poder determinar los valores sospechosos o anormales aplicando los Límites para esos casos.

EJEMPLO 19. Se toman dos muestras aleatorias de los ingresos diarios, en miles de pesos, de dos hoteles en la temporada de alza turística. Los datos aparecen en la siguiente tabla:

H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	72	75	56	74	66	78	75	70	68	67	70	86	69	73	70
2	56	55	54	53	57	56	55	57	54	51	55	57	58	52	56

A partir de la información puede construirse los Diagramas de Caja para cada una de las variables. Ordenando ambas variables en orden ascendente, la tabla de los datos quedaría:

H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	56	66	67	68	69	70	70	70	72	73	74	75	75	78	86
2	51	52	53	54	54	55	55	55	56	56	56	57	57	57	58

Para el caso del Hotel 1 puede determinarse que la Mediana es igual a 70, el Cuartil 25 es igual a 68 y el Cuartil 75 es igual a 75. Con esos valores puede calcularse lo siguiente:

$$RIC = C_{75} - C_{25} = 75 - 68 = 7$$

$$LIS = C_{25} - 1.5 * RIC = 68 - 10.5 = 57.5$$

$$LSS = C_{75} + 1.5 * RIC = 75 + 10.5 = 85.5$$

$$LIA = C_{25} - 3 * RIC = 68 - 21 = 48$$

$$LSA = C_{75} + 3 * RIC = 75 + 21 = 96$$

Nótese que el valor de 56 está dentro de los datos sospechosos y el 98 es un dato anormal.

Con la información anterior puede construirse un Diagrama de Caja, el que se muestra en el siguiente gráfico.

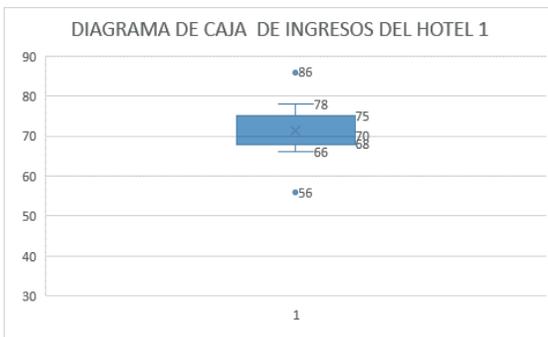


Gráfico 12. Diagrama de caja de los ingresos del hotel 1.

Como se puede apreciar, en este caso, los bigotes se han seleccionado como el menor y mayor valor de la muestra después que se eliminan los valores sospechosos, en este caso 56 y anormales, en este caso 86. Del gráfico puede interpretarse que los datos tienen asimetría con respecto a la Mediana, existiendo mayor cantidad de datos por encima que por debajo.

Más información brinda este diagrama, cuando se analizan simultáneamente varias muestras. En la figura de abajo se muestra los Diagramas de Cajas para los ingresos diarios de los hoteles 1 y 2.

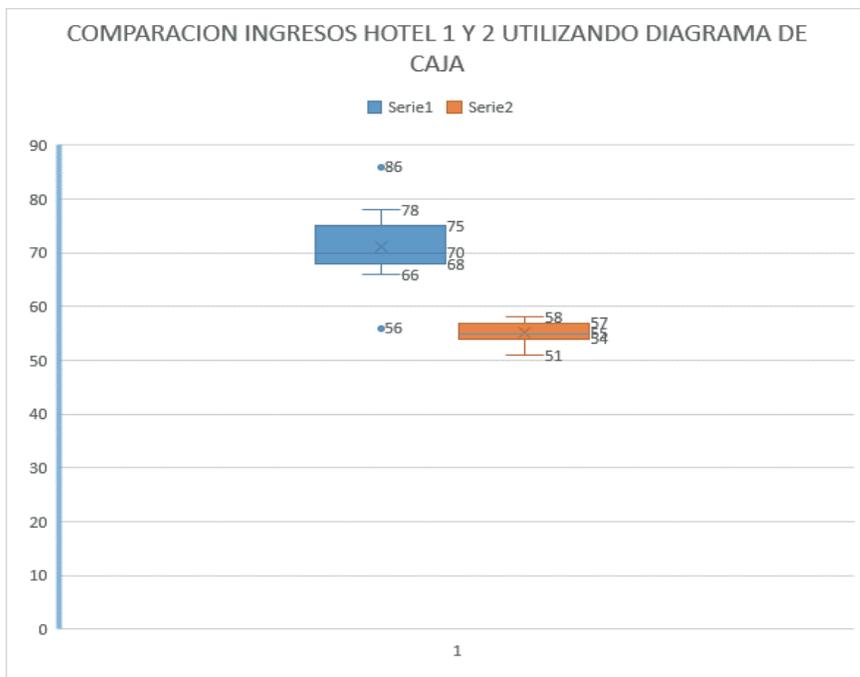


Gráfico 13. Comparación de los datos del hotel 1 y 2, a través de un diagrama de caja.

Puede observarse:

- La mediana del Hotel 1 (70) es ligeramente superior a la del Hotel 2 (55).
- La dispersión de los datos del Hotel 1 es mayor que la del Hotel 2, dado que el Rango Intercuartil es mayor.
- Los datos del Hotel 2 son más simétricos con respecto a la Mediana que los del Hotel 1.

- En el Hotel 1 aparece un valor sospechoso y otro anormal y no ocurre así con los datos del Hotel 2.

El EXCEL puede utilizarse para la construcción de Diagramas de Caja.

Para ello se colocan cada juego de datos en distintas columnas de una Hoja EXCEL. Se marca la columna a la cual se le quiere construir el Diagrama de Caja y se inserta entre los gráficos estadísticos que aparecen el Diagrama de Caja como se muestra en la Figura 75.

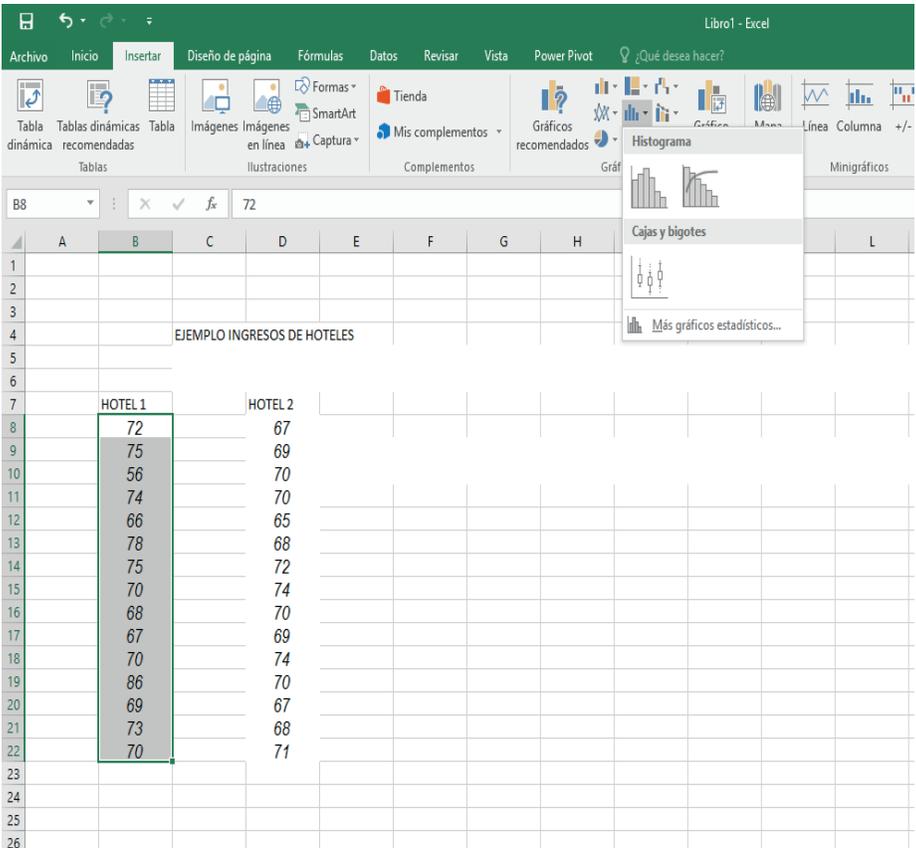


Figura 75. Hoja de Excel con los datos para la construcción de diagrama de caja.

Cuando se marca el Diagrama de Caja, aparece inmediatamente en la misma Hoja el gráfico correspondiente:

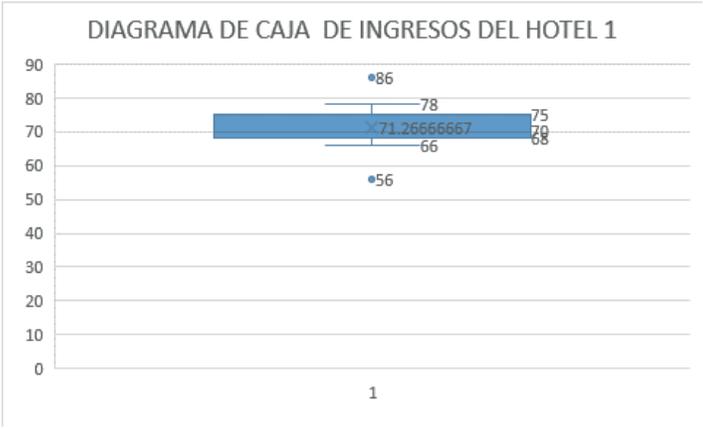


Gráfico 14. Diagrama de caja a través de Excel.

En este caso identifica la Mediana (70), los cuartiles Q_1 (68) y Q_3 (75) y el valor de los bigotes (66 y 78), así también como las observaciones anormales que deben revisarse (56 y 86). También incorpora el valor del Promedio (71.27).

Cuando se desea hacer comparaciones de juegos de datos utilizando el Diagrama de Caja se marcan cada columna de datos por separado y se inserta el Diagrama dentro de los gráficos estadísticos y enseguida aparecerá el Diagrama correspondiente, tal como se muestra en el gráfico 15:

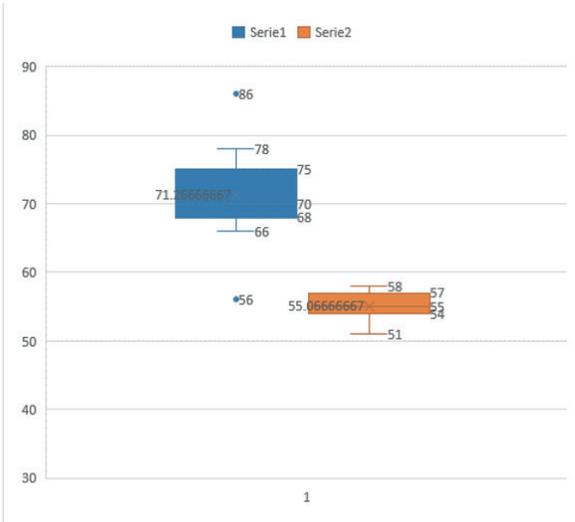


Gráfico 15. Construcción de dos series de datos utilizando Excel.

3.2. Problemas resueltos sobre medidas descriptivas

En esta sección se resolverán los problemas utilizando los software estadísticos, que se han estado explicando.

PR 13. La cantidad de turistas alojados en una instalación turística cada día, es una variable aleatoria. Se ha tomado una muestra de 10 días y la información de muestra en la siguiente tabla:

152	164	146	151	168	156	150	165	178	168
150	157	162	176	146	151	165	166	158	149
158	161	165	160	176	165	170	158	173	165
170	162	154	163	149	158	162	169	154	178
167	176	152	168	176	165	153	146	154	156
174	157	171	165	159	162	176	157	148	154
159	164	173	177	167	166	154	150	160	154
168	159	147	171	164	176	155	165	161	154
167	149	151	158	160	156	164	166	171	173
158	169	165	156	171	166	170	163	150	163

Con esos datos se quiere calcular:

- Promedio o media aritmética del número de personas alojadas en un día.
- Mediana
- Desviación Estándar.
- Los cuartiles 1 y 3.
- Rango intercuartil
- Índice de simetría.
- Índice de apuntalamiento o curtosis.
- Construir el Diagrama de Caja para estos datos.

SOLUCIÓN

Primero se resolverá el problema utilizando el EXCEL y posteriormente el SPSS.

Para ello se colocarán los datos en una hoja EXCEL, en la misma forma matricial en que aparece o en una sola columna. Para trabajar en una misma hoja todos los estadígrafos descriptivos, se puede preparar una hoja colocando en una secuencia de celdas los nombres de los que se deseen calcular, tal como se muestra en la Figura 76.

Prob resueltos estadígrafos - Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista Power Pivot ¿Qué desea hacer?

D23

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		DATOS DE TURISTAS ALOJADOS EN UN DÍA EN UNA INSTALACIÓN HOTELERA PR 1.13										
3												
4		152	164	146	151	168	156	150	165	178	168	
5		150	157	162	176	146	151	165	166	158	149	
6		158	161	165	160	176	165	170	158	173	165	
7		170	162	154	163	149	158	162	169	154	178	
8		167	176	152	168	176	165	153	146	154	156	
9		174	157	171	165	159	162	176	157	148	154	
10		159	164	173	177	167	166	154	150	160	154	
11		168	159	147	171	164	176	155	165	161	154	
12		167	149	151	158	160	156	164	166	171	173	
13		158	169	165	156	171	166	170	163	150	163	
14												
15												
16		PROMEDIO										
17		MEDIANA										
18		DESVIACIÓN ESTÁNDAR										
19		CUARTIL 1 (Q1)										
20		CUARTIL 3 (Q3)										
21		RANGO INTERCUARTIL										
22		INDICE DE SIMETRÍA										
23		INDICE DE CURTOSIS										
24												
25												
26												
27												
28												
29												
30												
31												

Figura 76. Hoja de Cálculo para la determinación de los estadígrafos descriptivos.

PR 1.14.

Con la Hoja EXCEL de esa forma, se selecciona una celda a la derecha del estadígrafo que se desea calcular y se busca las Funciones Estadísticas que proporciona el EXCEL la deseada para cada caso. Esto se muestra en la Figura 77.

Prob resultados estadígrafos - Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista Power Pivot ¿Qué desea hacer?

D16

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		DATOS DE TURISTAS ALOJADOS EN UN DÍA EN UNA INSTALACIÓN HOTELERA PR 1.13										
3												
4		152	164	146	151	168	156	150	185	178	188	
5		150	157	162	176	146	151	165	166	158	149	
6		158	161	165	160	176	185	170	158	173	165	
7		170	162	154	163	149	158	162	169	154	178	
8		167	176	152	168	176	185	153	146	154	156	
9		174	157	171	165	159	162	176	157	148	154	
10		159	164	173	177	167	166	154	150	160	154	
11		168	159	147	171	164	176	155	165	161	154	
12		167	149	151	158	160	156	164	166	171	173	
13		158	169	165	156	171	166	170	163	150	163	
14												
15												
16		PROMEDIO	=									
17		MEDIANA										
18		DESVIACIÓN ESTÁNDAR										
19		CUARTIL 1 (Q1)										
20		CUARTIL 3 (Q3)										
21		RANGO INTERCUARTIL										
22		INDICE DE SIMETRÍA										
23		INDICE DE CURTOSIS										
24												
25												
26												
27												
28												
29												
30												
31												
32												
33												

Insertar función

Buscar una función:

Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir

O seleccionar una categoría: Estadísticas

Seleccionar una función:

- PERMUTACIONES
- PERMUTACIONES.A
- POISSON.DIST
- PROBABILIDAD
- PROMEDIO**
- PROMEDIO.SI
- PROMEDIO.SI.CONJUNTO

PROMEDIO(número1,número2,...)

Devuelve el promedio (media aritmética) de los argumentos, los cuales pueden ser números, nombres, matrices o referencias que contengan números.

[Ayuda sobre esta función](#)

Aceptar Cancelar

Figura 77. Ventana para la selección de las funciones estadísticas del Excel.

Para calcular el Promedio, se selecciona la función PROMEDIO y se abrirá una ventana para entrar la información necesaria; en ella se colocarán los datos del problema, marcando el rango de las celdas donde están los datos, tal como se muestra en la Figura 78.

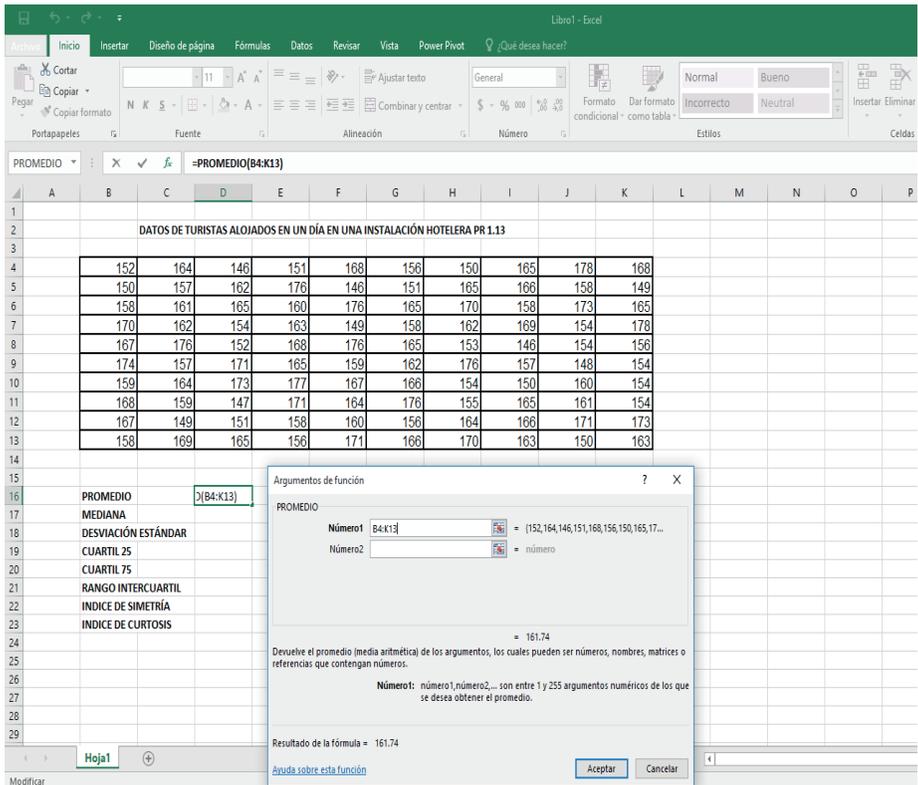


Figura 78. Ventana Excel para el cálculo de la media aritmética o el promedio.

En la misma ventana aparece el valor del promedio (161.74) y cuando se marca Aceptar aparecerá este valor en la celda seleccionada para el valor del promedio.

De forma similar se trabajará para el cálculo del resto de los estadígrafos, utilizando las Funciones Estadísticas del EXCEL, MEDIANA, DESVEST.M, CUARTIL.INC, COEFICIENTE.ASIMETRÍA Y CURTOSIS. El Rango Intercuartil se calcula directamente sin función.

b) Mediana

La ventana de Excel para el cálculo de la mediana se muestra en la figura 79.

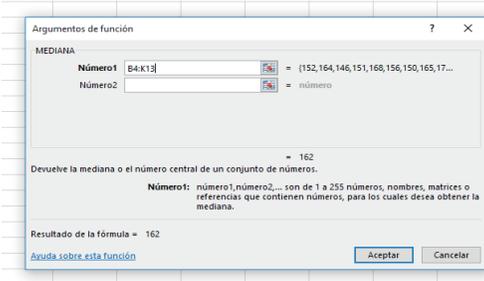


Figura 79. La ventana de Excel para el cálculo de la mediana.

El valor de la Mediana puede leerse directamente en la ventana y en este caso es 162 turistas cada día.

c) Desviación típica o estándar

Se selecciona en las Funciones Estadísticas del EXCEL la función DESVEST.M y se procede de forma similar a lo realizado anteriormente, colocando los datos en la casilla correspondiente de la ventana para esta función y en la misma ventana aparece el valor de 8.51 turistas por día para este estadígrafo, tal como se muestra en la Figura 80.

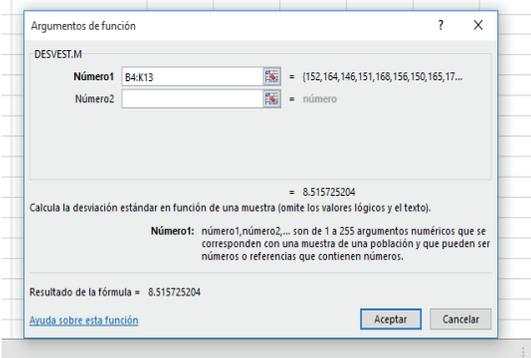


Figura 80. Ventana de Excel para el cálculo de la desviación estándar.

d) Para el cálculo de los Cuartiles, utilizando el EXCEL, se utiliza la Función Estadística CUARTIL.EXC y una vez seleccionada aparece la ventana correspondiente para esta función. Se introduce los datos de la matriz de datos del problema y en la celda de Cuartil se calcula primero el Cuartil 1 y después el Cuartil 3. Ambas ventanas con el valor de los cuartiles se muestran en la Figura 81.

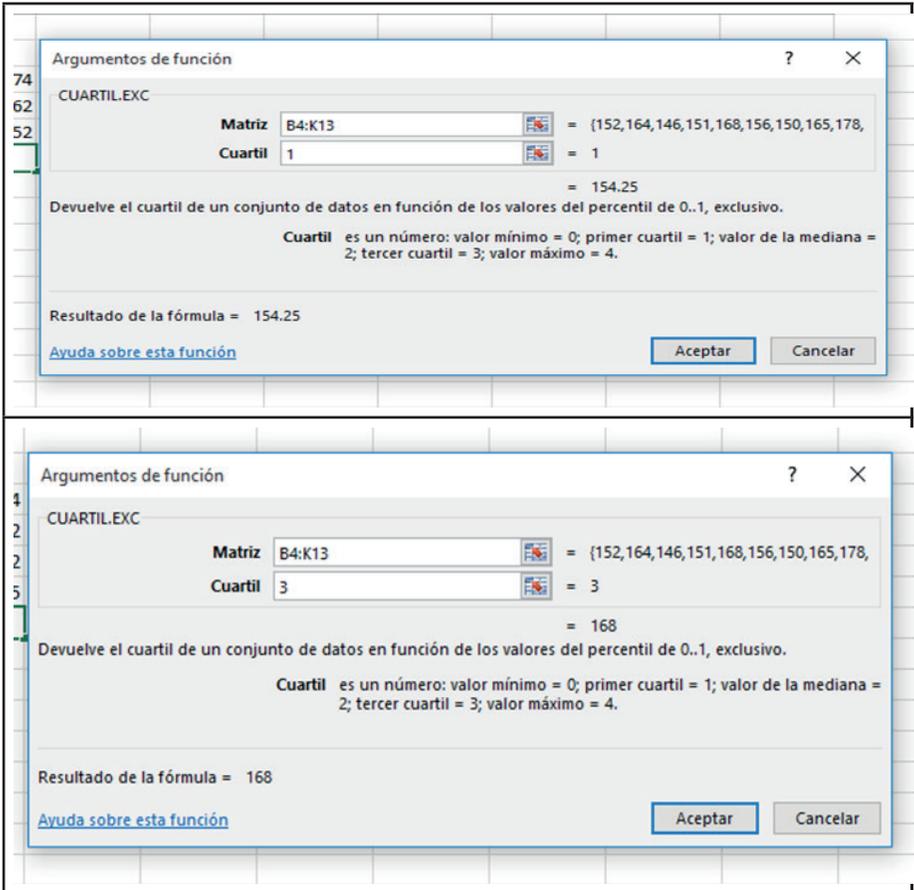


Figura 81. Ventanas para el cálculo de los cuartiles.

Los valores de ambos cuartiles son $Q_1 = 154.25$ y $Q_3 = 168$

e) El Rango Intercuartil por su sencillez, puede calcularse manualmente a partir del valor de Q_1 y Q_3 .

$$RIC = Q_3 - Q_1 = 168 - 154.25 = 13.75 \text{ turistas por día}$$

Antes de calcular el Índice de Asimetría y de Curtosis, se muestra en el gráfico 16 el Histograma de Frecuencia Absoluta para los datos del problema, lo que permite una mejor comprensión de estos índices.

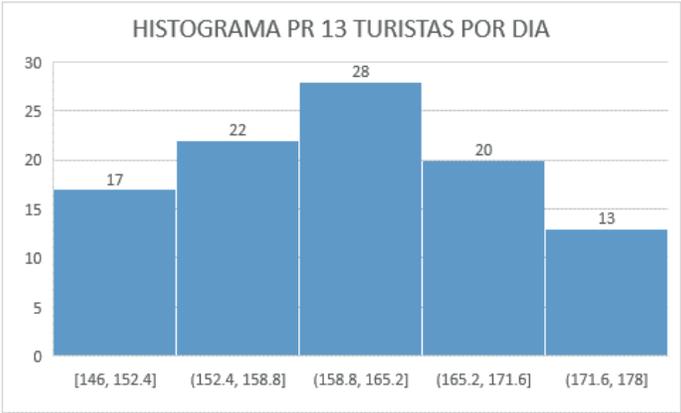


Gráfico 16. Histograma del PR 13.

f) Índice de simetría

Para calcular el Índice de Asimetría, se selecciona la Función Estadística de EXCEL COEFICIENTE DE ASIMETRIA, apareciendo la ventana que permite introducir los datos del problema, la que se muestra en la Figura 82.

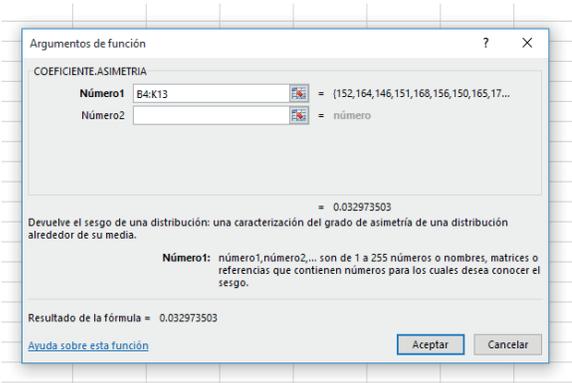


Figura 82. Ventana de Excel para el cálculo del coeficiente de asimetría.

Directamente en la ventana puede verse que este índice tiene un valor de 0.03, bastante cercano al valor 0, por lo que puede concluirse que la distribución de los datos es simétrica con respecto a su media aritmética o promedio, lo que puede visualizarse en el Histograma de Frecuencia Absoluta mostrado anteriormente.

g) Índice o coeficiente de Curtosis.

De forma similar a lo realizado hasta el momento en las Funciones Estadísticas del EXCEL la función CURTOSIS y se introduce en la ventana que aparece los datos del problema, tal como se muestra en la Figura 83.

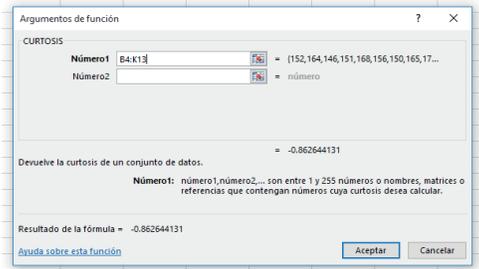


Figura 83. Ventana de Excel para el cálculo del coeficiente de curtosis.

Nótese que el valor del Índice o Coeficiente de Curtosis es de -0.8626, por lo que puede concluirse que la forma de la distribución de los datos del problema es platicúrtica, o sea, una distribución achatada con relación a su media, lo que puede comprobarse con el Histograma mostrado.

h) Construir el Diagrama de Caja.

Para construir el diagrama de caja de este problema, hay que poner los datos en una fila o columna y no en la forma matricial en que se ha trabajado hasta el momento. Se selecciona esta fila o columna y en Insertar se selecciona el gráfico estadístico asociado al Diagrama de Caja, lo que se muestra en la Figura 84.



Figura 84. Hoja de Excel para la construcción del diagrama de caja.

De forma inmediata aparece el Diagrama de Caja para los datos seleccionados y usando el editor se adiciona el título y las etiquetas de los datos, quedando de la forma que se muestra en el gráfico 17.

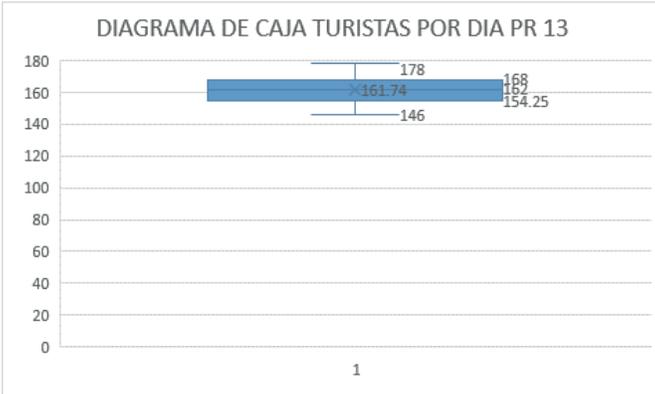


Gráfico 17. Diagrama de caja para el PR 13.

Del gráfico puede deducirse que hay bastante concentración de los datos con relación a su Mediana (162) y no hay datos sospechosos, ni con valores extremos.

Finalmente se muestra en la Figura 854 la Hoja EXCEL que resume todos los resultados de este problema.

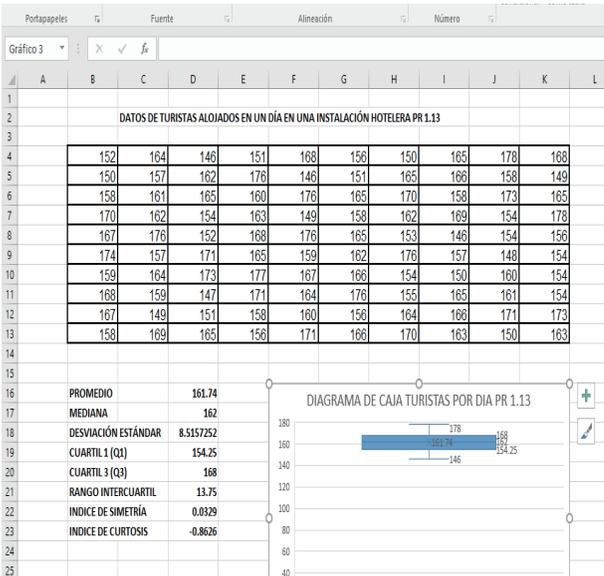


Figura 85. Hoja de Excel con los estadígrafos calculados.

PR 14. En el PR 6, relativo a la nota de una prueba diagnóstico de los conocimientos en Matemática de los alumnos que ingresaban a una universidad, se brindaban los siguientes datos:

65	41	54	63	86	67	93	71	63	57	77	59
69	66	74	56	75	61	95	84	73	73	81	58
66	57	66	74	58	73	99	39	70	57	68	59
75	41	57	48	82	67	98	91	60	73	59	65
77	56	44	75	90	95	48	53	91	72	63	81

Con los datos anteriores calcule:

- Promedio o media aritmética del número de personas alojadas en un día.
- Mediana
- Varianza y Desviación Estándar.
- Los cuartiles 1 y 3
- Rango intercuartil
- Índice de simetría.
- Índice de apuntalamiento o curtosis.
- Construir el Diagrama de Caja para estos datos.

SOLUCIÓN

Similar al problema anterior, se resolverá este aplicando las Funciones Estadísticas del EXCEL.

- Calculo del promedio o media aritmética.

Se trabajará desde un inicio con los datos colocados en una columna de la Hoja EXCEL y para el cálculo del promedio se selecciona la Función PROMEDIO dentro de las Funciones Estadísticas del EXCEL y aparecerá la ventana que se muestra en la Figura 86.

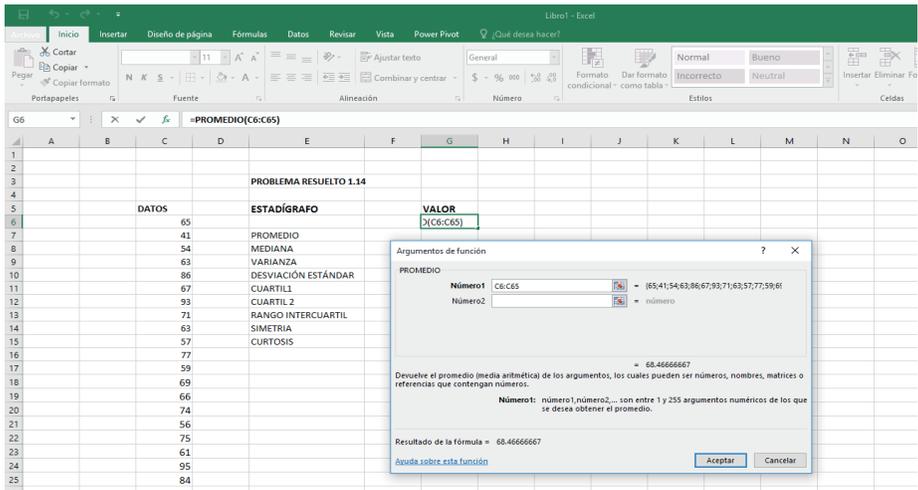


Figura 86. Hoja de Excel para cálculo del promedio.

Por lo que el valor promedio aproximado se puede ver en la propia ventana y es de 68.46 puntos.

Para una mejor interpretación y discusión de los próximos estadísticos, se ha construido el Histograma de Frecuencia Absoluta para estos datos, el que se muestra en el gráfico 18.

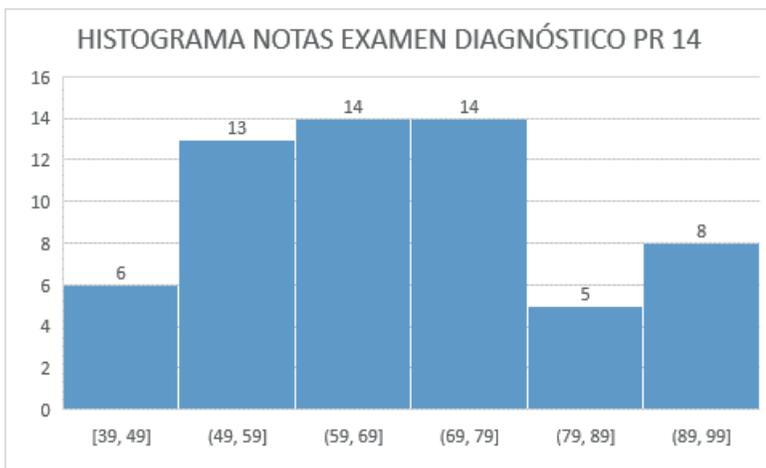


Gráfico 18. Histograma de Frecuencia Absoluta.

b) De forma similar como se calculó el promedio, se calcularán los otros estadígrafos. La ventana para la Mediana, se muestra en la Figura 87.

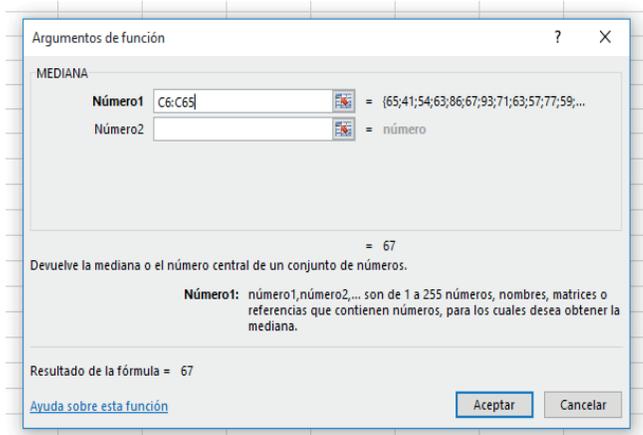


Figura 87. Ventana de Excel para el cálculo de la mediana.

La Mediana tiene un valor de 67 puntos, ligeramente inferior al promedio, el que está influenciado por valores más dispersos hacia la derecha de la Mediana. En el Histograma puede verse que la Mediana está dentro del intervalo (59,69). A la derecha de este intervalo hay 27 valores en tres intervalos, mientras que a la izquierda hay 19 en dos intervalos, de ahí que el Promedio sea ligeramente superior a la Mediana.

c) Las ventanas para el cálculo de la Varianza y la Desviación Estándar se muestran en la Figura 88 y 89.

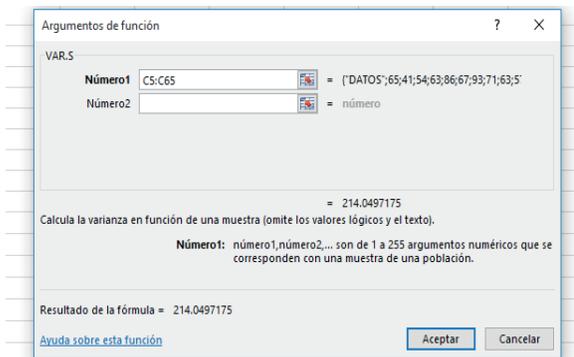


Figura 88. Ventana de Excel para el cálculo de la varianza.

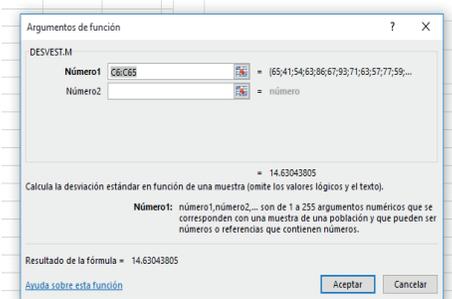


Figura 89. Ventana de Excel para el cálculo de la desviación estándar o típica.

La Varianza tiene un valor aproximado de 214.05 y la Desviación Estándar de 14.63. Calculando el Coeficiente de Variación:

$$* 100 = (14.63/68.46)*100 = 21.4\%$$

Lo que indica un valor moderado de dispersión con relación al promedio. Esto puede ser verificado en el Histograma.

d) El cálculo de los Cuartiles 1 y 3 se realiza utilizando la Función CUARTIL. EXC del EXCEL. Una vez que se selecciona esta función se abre en la Hoja EXCEL la correspondiente ventana y se introduce la matriz de los datos que aparece en la columna C, en la celda del Cuartil se pone el número del cuartil a calcular y directamente aparece en estas ventanas los valores de los cuartiles correspondientes, tal como se muestra en la Figura 90 y 91.

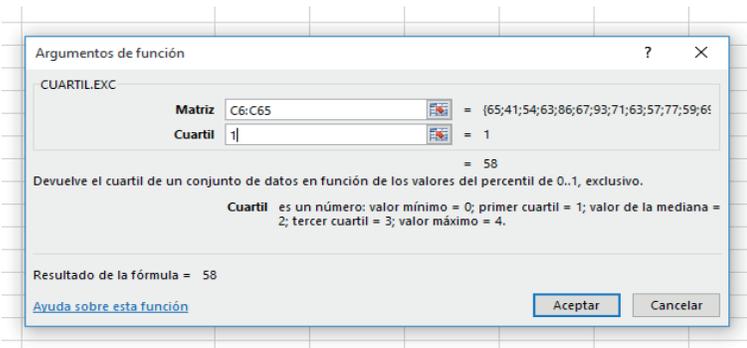


Figura 90. Ventana de Excel para el cálculo del cuartil 1.

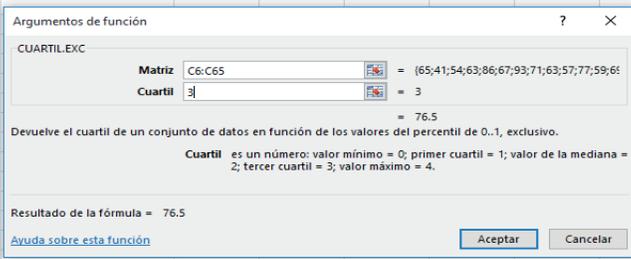


Figura 91. Ventana de Excel para el cálculo del cuartil 3.

Y los valores de los Cuartiles serán $Q_1 = 58$ puntos y $Q_3 = 76.5$

e) El Rango Inter cuartil se calculará manualmente utilizando los datos de los Cuartiles y este es:

$$RIC = Q_3 - Q_1 = 76.5 - 58 = 18.5 \text{ puntos}$$

f) Para calcular el Coeficiente o Índice de Asimetría se utiliza la función correspondiente del EXCEL y una vez seleccionada aparecerá en la Hoja EXCEL la ventana correspondiente, la que se muestra en la Figura 92, donde se ha colocado los datos para el cálculo de este estadígrafo para el problema.

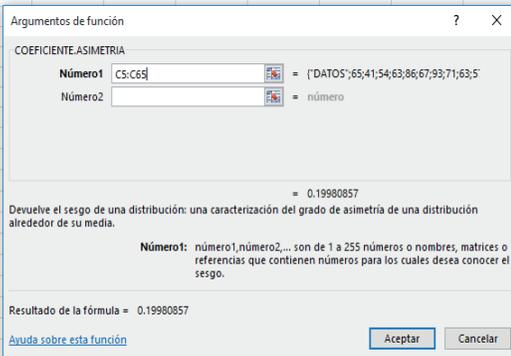


Figura 92. Ventana de Excel para el cálculo para el coeficiente de asimetría.

Su valor es de 0.1998, lo que implica una ligera asimetría a la derecha, lo que puede verificarse en el Histograma.

g) La ventana para el cálculo de la Curtosis se muestra en la Figura 93.

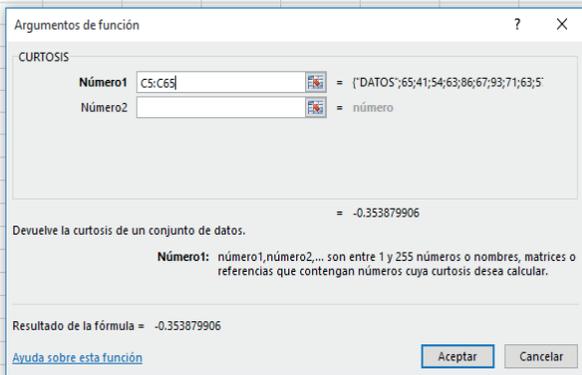


Figura 93. Ventana de Excel para el cálculo de la curtosis.

El valor del Índice de Curtosis es de aproximadamente - 0.3538 y al ser negativa indica que es platicurtica, o sea ligeramente achatada. Note que en el Histograma de Frecuencias Absolutas mostrado anteriormente hay 3 intervalos que prácticamente tienen igual frecuencia lo que contribuye a esta característica de forma.

El Diagrama de Caja para este conjunto de datos, se muestra en el gráfico 19.

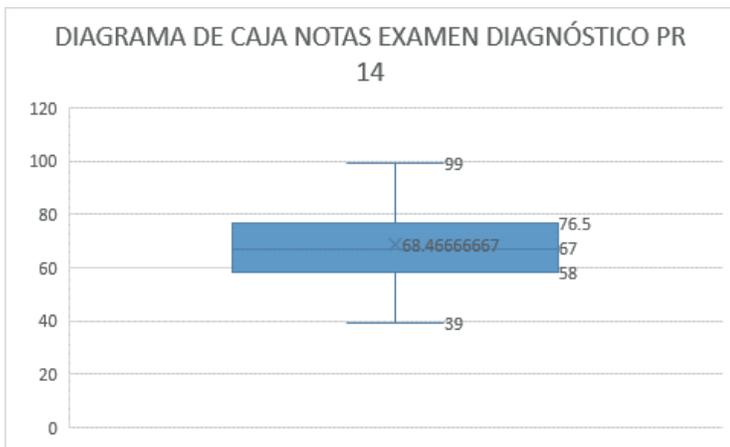


Gráfico 19. Diagrama de caja del ejercicio notas de examen de diagnóstico.

Como puede observarse hay cierta simetría con respecto a la Mediana, ya que el límite superior de la caja (76.5) está a 6.5 unidades por arriba de la

media y el punto inferior (58) a 9 unidades. La dispersión de los datos dada entre el valor más pequeño (39) y el mayor (99) es de 60 unidades. Si se compara con el problema anterior, los datos de este problema tienen mayor variabilidad, ya que en el ejemplo anterior este valor era de 32 unidades.

Finalmente se muestra la Hija EXCEL con el resumen de todo el problema resuelto en la Figura 94.

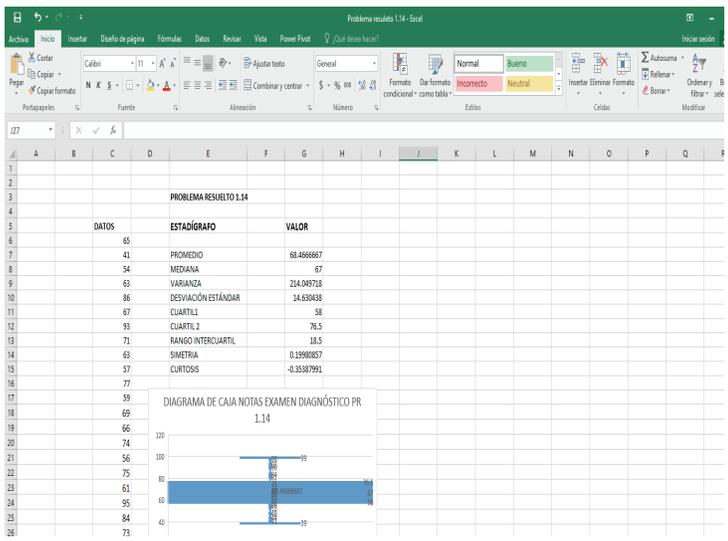


Figura 94: Hoja de Excel con los resultados finales de los estadígrafos del PR. 14.

PR 15. El consumo de combustible (fuel oil), en litros por semana, en una microempresa dedicada a la elaboración de alimentos es aleatoria y se han recopilado los datos para semanas de este combustible y estos se muestran en la siguiente tabla:

7638	7420	8825	8584	7918	6903	5768	5771	9114	5506
7990	7621	8816	8774	7670	8260	5887	5557	4450	3150
4780	5730	5720	5760	5930	4560	5266	4780	5280	5630

Con esto datos, calcule>

a) Promedio o media aritmética de la cantidad de litros consumidas semanalmente en esa microempresa.

b) Mediana

- c) Varianza y Desviación Estándar.
- d) Los cuartiles 1 y 3
- e) Rango intercuartil
- f) Índice de simetría.
- g) Índice de apuntalamiento o curtosis.
- h) Construir el Diagrama de Caja para estos datos.

SOLUCIÓN

Como en los problemas resueltos anteriormente, es útil construir una tabla inicial donde se coloquen los datos en una columna o en una matriz y se agregue los nombres de los estadígrafos que se desean calcular.

a) Para el cálculo del promedio o media aritmética se selecciona la función estadística PROMEDIO de EXCEL y en la celda de Números se pone el intervalo de celdas donde aparecen los datos. Esto se muestra en la Figura 95.

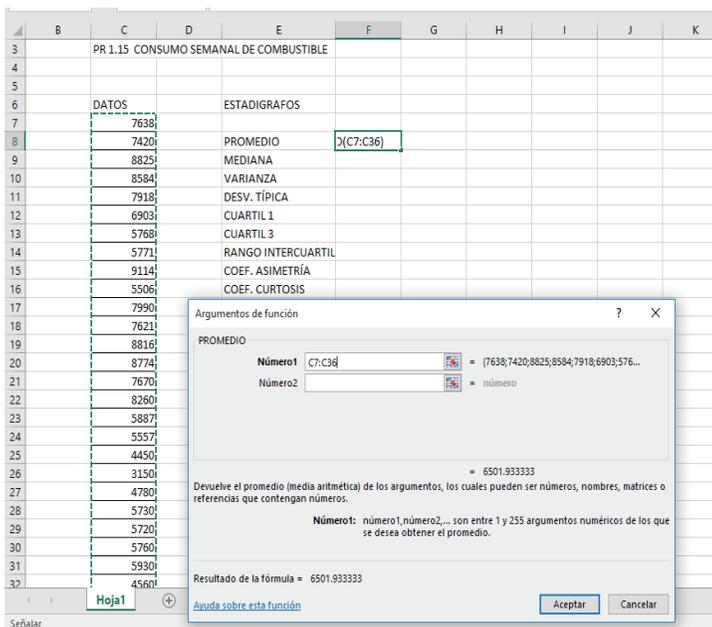


Figura 95. Hoja de Excel para el cálculo del promedio de PR. 15.

Cuando se marca aceptar aparecería el valor promedio en la casilla seleccionada de la Hoja EXCEL, aunque directamente puede leerse el valor en la ventana de la función y este es 6505.93 litros por semana.

De forma similar pueden encontrarse los valores de los estadígrafos, Mediana, Varianza, Desviación Típica o Estándar y los Cuartiles 1 y 3. Las ventanas para el cálculo de estos estadígrafos se muestran en las Figuras de la 96 a la 100.

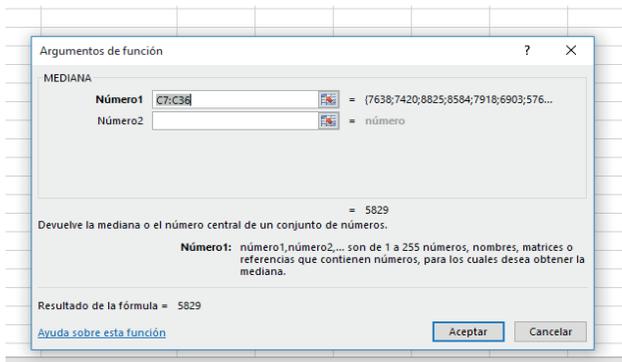


Figura 96. Ventana de Excel para el cálculo de la mediana del PR. 15.

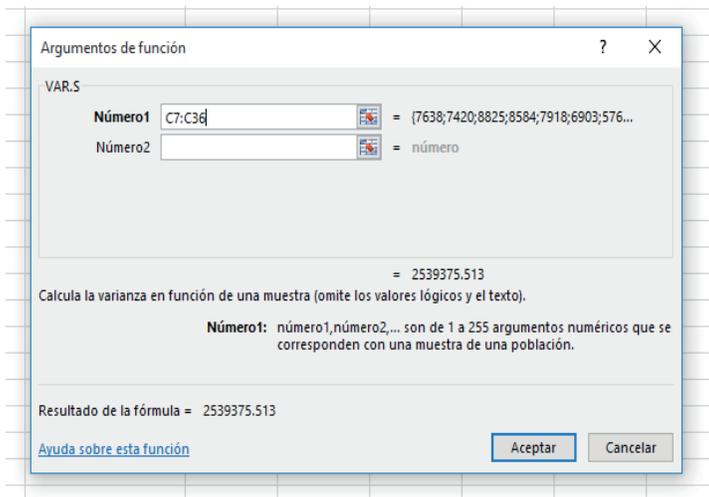


Figura 97. Ventana de Excel para el cálculo de la varianza del PR. 15.

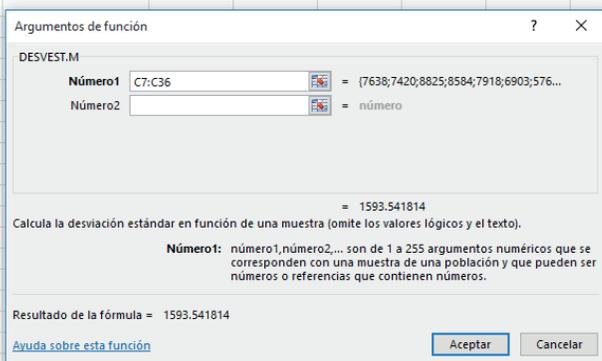


Figura 98. Ventana de Excel para el cálculo de la desviación estándar del PR. 15.

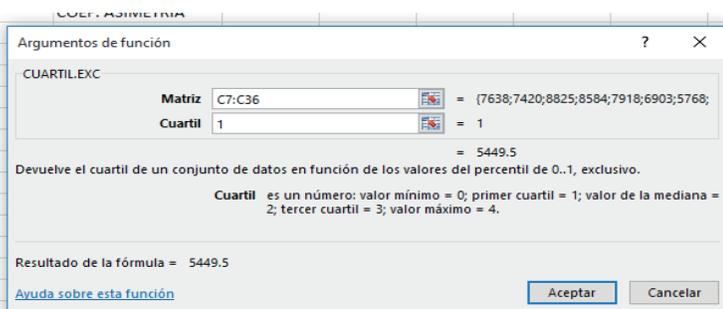


Figura 99. Ventana de Excel para el cálculo del cuartil 1 del PR. 15.

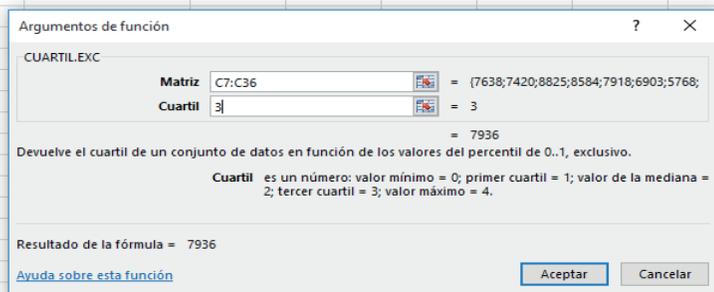


Figura 100. Ventana de Excel para el cálculo el cuartil 3 del PR. 15.

El Rango Intercuartil, puede calcularse por EXCEL o directamente

$$RIC = Q_3 - Q_1 = 7936 - 5449.5 = 2486.5 \text{ litros por semana}$$

Finalmente se calcula los estadígrafos de forma utilizando el EXCEL, cuyas ventanas y valores aparece en la Figura 101 y 102.

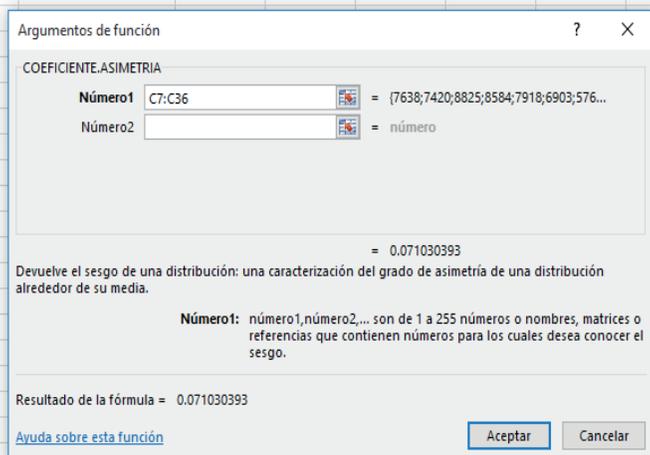


Figura 101. Ventana de Excel para el cálculo del coeficiente de asimetría del PR. 15.

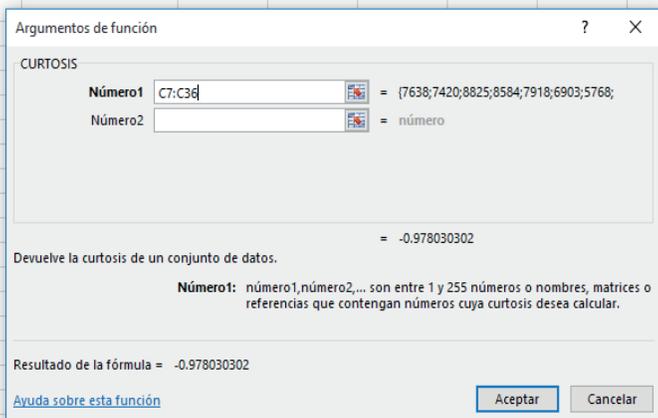


Figura 102 Ventana de Excel para el cálculo del coeficiente de curtosis del PR. 15.

Para una mejor interpretación de los resultados, se construye un Histograma de Frecuencia Absoluto y el Diagrama de Caja de este problema, los que se muestran en el gráfico 20.

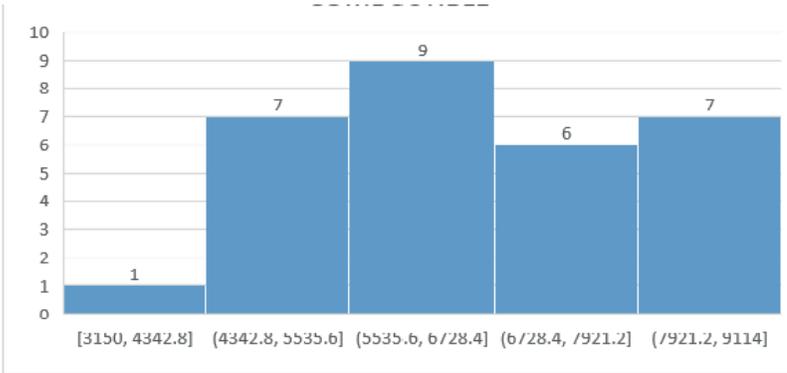


Gráfico 20. Histograma de Frecuencia Absoluta del PR. 15.

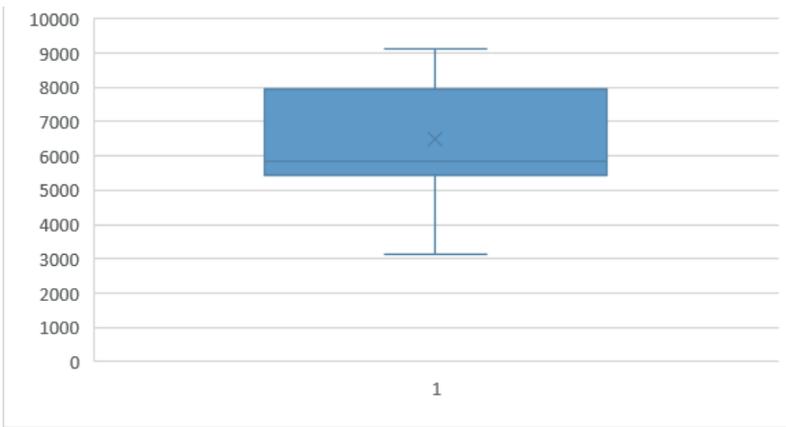


Gráfico 21. Diagrama de caja del PR.15.

Finalmente, la Tabla con los resultados en la Hoja EXCEL se muestra en la figura 103.

PR 1.15 CONSUMO SEMANAL DE COMBUSTIBLE			
DATOS	ESTADIGRAFOS		
7638			
7420	PROMEDIO		6501.93333
8825	MEDIANA		5829
8584	VARIANZA		2539975.51
7918	DEV. TÍPICA		1593.54181
6903	CUARTIL 1		5449.5
5788	CUARTIL 3		7996
5771	RANGO INTERCUARTIL		2486.5
9114	COEF. ASIMETRÍA		0.07103039
5506	COEF. CURTOSIS		-0.9780303
7990			
7621			

Figura 103. Hoja de Excel para el cálculo de los estadígrafos del PR. 15.

Un análisis integral de estos estadígrafos indica:

Hay una diferencia notable entre el Promedio (6501.93) y la Mediana (5829). Note que la Desviación Típica tiene un valor alto de 1593.54, lo que implicaría un Coeficiente de variación de casi un 25%, esto es hay dispersión en los datos, y los valores extremos pueden estar influyendo en esta diferencia entre estos dos estadígrafos de Medida Central. Esto implica la necesidad de estudiar las causas de esta dispersión en el consumo de combustible de cada semana en esa microempresa.

El Coeficiente de Asimetría (0.07) es positivo, aunque cercano a cero. Puede concluirse que hay una ligera asimetría hacia la derecha con relación al valor del Promedio. Esto puede verificarse observando el Histograma de Frecuencias Absolutas, donde puede verse que se agrupan más observaciones a la derecha del Promedio, que estaría casi en el punto medio del intervalo (5535.6 ; 6728.4). En el Diagrama de Caja también puede observarse que hay asimetría con relación a la Mediana.

Finalmente, el valor del Coeficiente de Curtosis es negativo (-0.9780), por lo que puede afirmarse que la distribución de los datos es platicúrtica, o sea achatada, lo que puede comprobarse al observar el Histograma de Frecuencias.

PR 16. Una firma consultora, está realizando un estudio sobre el Clima Laboral de una empresa y ha elaborado una encuesta con 15 preguntas sobre la temática, para ser aplicada a obreros y técnicos los que responderán a cada afirmación en base a la siguiente escala.

- Siempre (5)
- Muchas veces (4)
- Algunas veces (3)
- Pocas veces (2)
- Nunca (1)

Se ha utilizado una escala ordinal para codificar esas respuestas, la que se muestra entre paréntesis.

Se ha aplicado la encuesta a 15 técnicos y 30 trabajadores y la respuesta a la afirmación: “Con frecuencia se piden y se ponen en práctica ideas de los subordinados”, e muestra en la siguiente tabla:

TÉCNICOS	3	3	3	3	2	3	3	3	3	4	3	2	4	3	4
OBREROS	2	2	3	2	1	2	2	1	3	1	1	2	3	2	1
	1	2	2	1	2	3	2	2	3	2	2	1	1	3	2

Procese estos datos utilizando los estadígrafos descriptivos y analice la respuesta.

SOLUCIÓN

En los estudios de administración es bastante frecuente que se utilice las encuestas con escalas ordinales para estudiar distintos problemas. Cuando la escala es ordinal y de pocos valores, en este caso 5 valores, es conveniente realizar los análisis, utilizando como medida de tendencia central la mediana y los estadígrafos de posición. Como en este caso se han encuestado a dos tipos de empleados, técnicos y obreros, resulta beneficioso comparar estos valore utilizando los Diagramas de Caja y para ello se utilizarán las herramientas del EXCEL.

El primer paso es construir una Hoja EXCEL donde se sitúen los datos del problema y los estadígrafos que se desean calcular, tal como se muestra en la Figura 104.

PR1.17

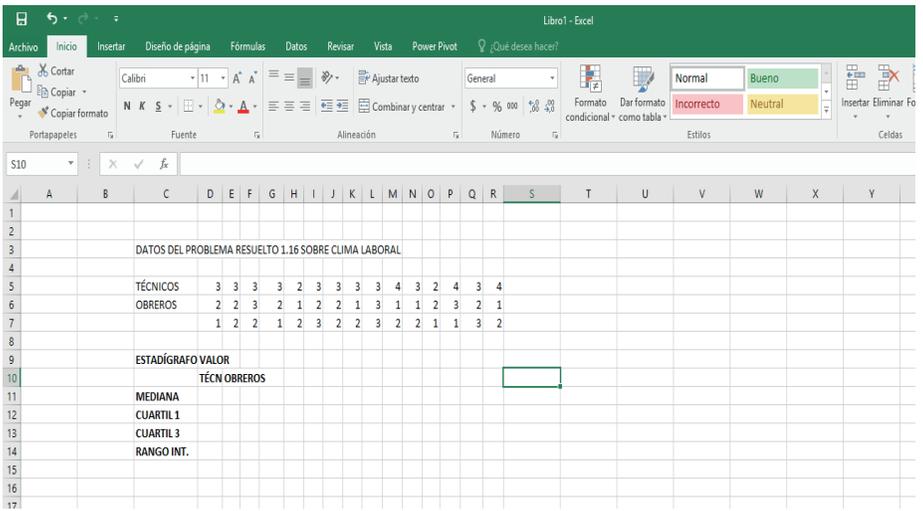


Figura 104. Hoja de Excel para solución del PR. 16.

Posteriormente utilizando las funciones estadísticas del EXCEL se pasa a calcular cada uno de los estadígrafos. Para el caso de la Mediana se muestran las ventanas para ambos casos en la Figura 105 y 106.

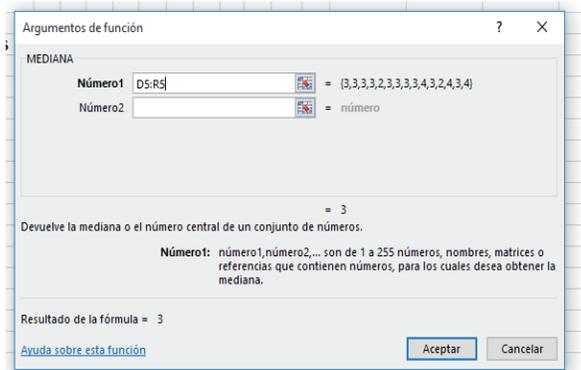


Figura 105. Ventana de Excel para el cálculo de la mediana de los técnicos del PR. 16.

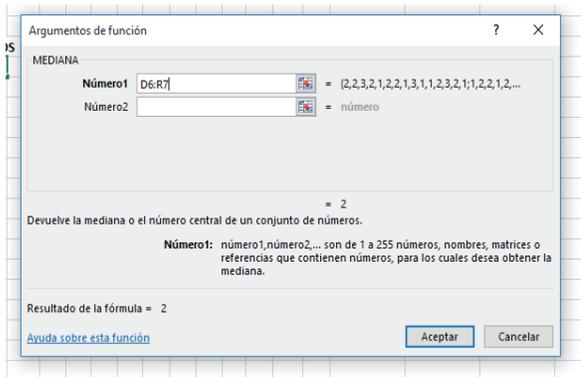


Figura 106. Ventana de Excel para el cálculo de la mediana de los obreros del PR. 16.

Con la misma filosofía de trabajo hasta el momento, se calcularían los cuartiles 1 y 3 para cada tipo de empleado. Para el caso de los obreros se muestran las ventanas en la Figura 107 y 108.

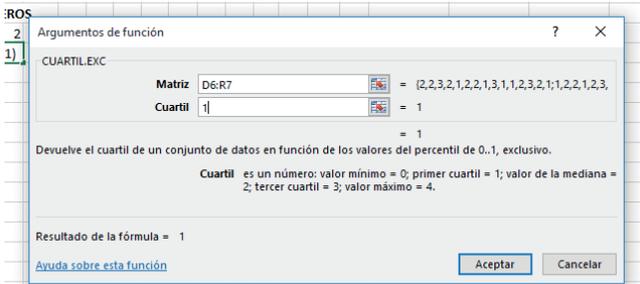


Figura 107. Ventana de Excel para el cálculo del cuartil 1 de los obreros del PR. 16.

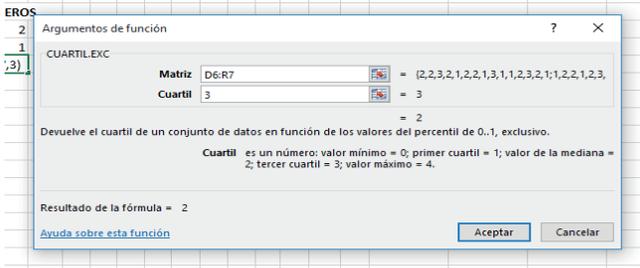


Figura 108. Ventana de Excel para el cálculo del cuartil 3 de los obreros del PR. 16.

La tabla con los resultados finales es:

Tabla 26. Estadísticos del PR. 16. Ecuador, 2018.

ESTADÍSTICO	VALOR		
	TÉCNICOS	OBREROS	
MEDIANA	3	2	
CUARTIL 1	3	1	
CUARTIL 3	3	2	
RANGO INT.	0	1	

Fuente: Datos ficticios.

De la tabla anterior puede interpretarse que hay diferencia en la mediana entre la apreciación que tienen los obreros y los técnicos en la respuesta a si se ponen en práctica las ideas de los subordinados. Los técnicos aprecian que algunas veces (3), mientras los obreros aprecian pocas veces (2). También puede apreciarse que en el 50% de los datos centrales, no hay dispersión en

la respuesta de los técnicos (rango intercuartil igual a cero), mientras que en los obreros los datos tienen mayor dispersión (RI igual a uno).

Para un análisis visual gráfico de lo expresado anteriormente se construye los Diagramas de Cajas para ambos tipos de empleados. Para ello en la Hoja EXCEL del problema, los datos de cada variable (técnicos y obreros) se colocan en una sola fila o columna. El diagrama resultante se muestra en el gráfico 22.

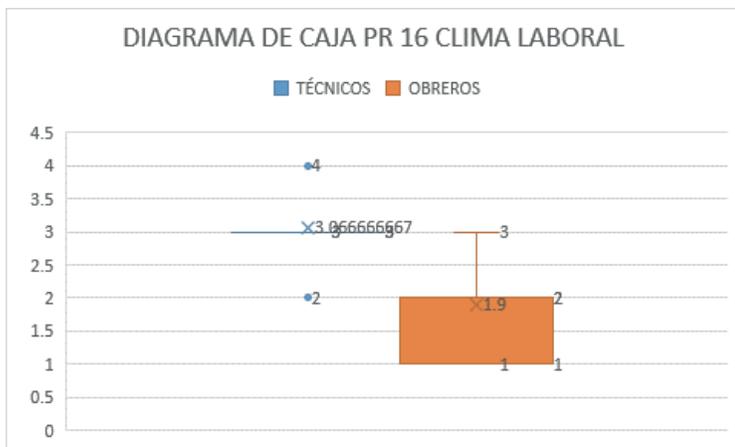


Gráfico 22. Diagrama de caja del PR. 16.

Se puede apreciar, como se discutió anteriormente, en el caso de los técnicos, la caja se representa por una línea sobre el valor de la mediana que es igual a 3 y se muestra que hay valores fuera de la caja con valor de 2 y 4. Por su parte en los obreros la mediana coincide con el límite inferior de la caja (1), pero hay una dispersión de 1, dado por la diferencia entre ambos lados de la caja. Fuera de la caja aparecen valores de 3. En la gráfica también aparecen los valores promedios.

Se puede concluir que hay distinta apreciación sobre la afirmación que se pide responder de los técnicos y los obreros de esta empresa, siendo mejor la de los técnicos que la de los obreros.

PR 17. Problema para el uso de la moda.

3.3. Valoración de los estadísticos descriptivos con el uso del SPSS

En la actualidad las entidades, Universidades, Empresas, Hospitales, entre otros tienen incorporado en sus procesos de análisis de datos el paquete

estadístico SPSS por las ventajas que este ofrece. El siguiente apartado tiene como objetivo analizar las salidas que ofrece el SPSS de los estadísticos descriptivos antes estudiados. Serán utilizados los ejemplos anteriormente valorados por el Excel.

EJEMPLO 20. Para ejemplificar se tomarán los 50 datos del ejemplo 3, del tiempo de atención en la Caja de un banco:

Una vez introducidos los datos en el SPSS, (para los que comienzan con el SPSS, pueden utilizar el libro Análisis exploratorio de datos con SPSS) se sigue el siguiente recorrido:

Ir a la figura 109

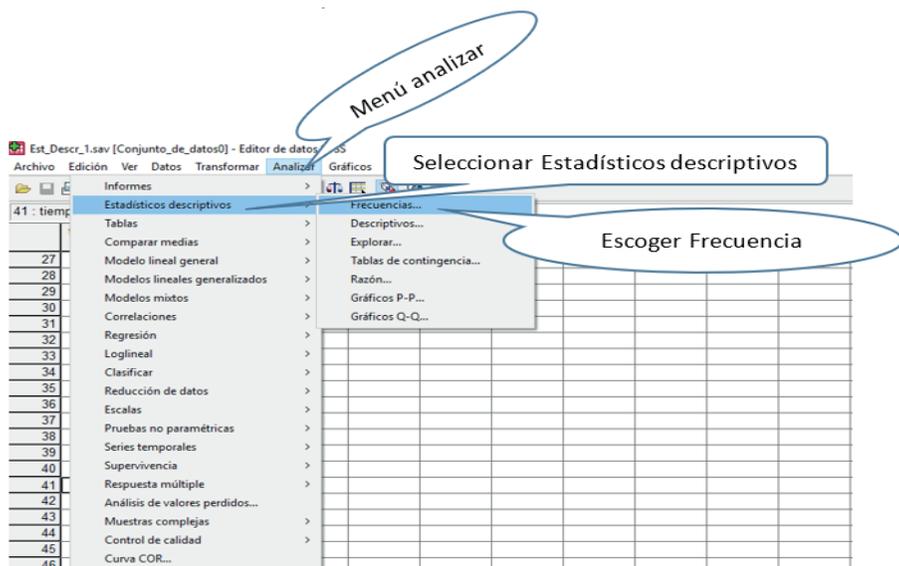


Figura 109. Ventana para la selección de los descriptivos de una variable de razón.

Se muestra la figura 109 el cómo se debe realizar el proceso para el cálculo de los estadísticos de una variable cualquiera que sea su medida, en el caso que ocupa, de razón.

En la figura 110 se muestra la venta generada por el SPSS al seleccionar “Frecuencia”,

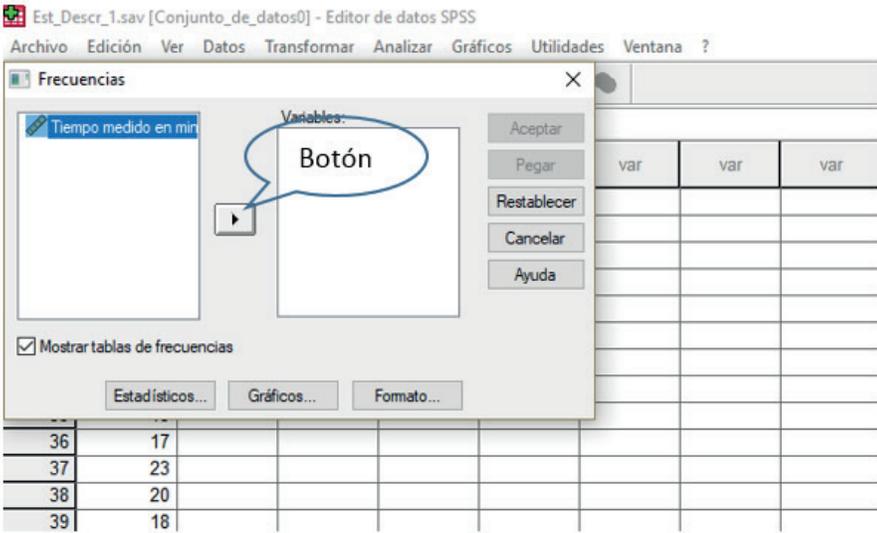


Figura 110. Ventana generada por el SPSS al seleccionar “Frecuencia, en “estadístico descriptivo” del panel Analizar.

En la figura 110 se escoge la variable que a procesar, en este ejemplo, solo hay una variable “tiempo de espera en caja de banco”, este proceso se puede realizar de varias formas:

- se hace opresión de doble clic izquierdo encima de la variable
- se marca la variable y se ejecuta el botón que se encuentra en el centro
- se arrastra la variable al cuadro “variables”

Una vez seleccionada la variable a procesar y oprimir “aceptar”, se genera la ventana de la figura 111.

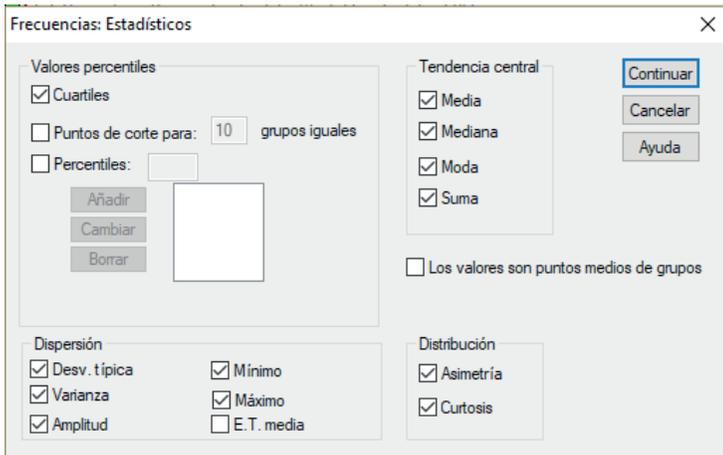


Figura 111. Ventana de Estadísticos descriptivos de la variable “Tiempo de espera en caja de banco”.

En esta figura 111 se observan cuatro tipos de estadísticos observados:

- Valores percentiles
- Tendencia central
- Dispersión
- Distribución

En el ejemplo que se está tratando, debido a que la variable es de razón, se ha podido seleccionar todas las opciones de los estadísticos, esta es una de las ventajas, a la hora de hacer un análisis de datos en SPSS, respecto al Excel, es decir, en una sola venta se generan todos los estadísticos descriptivos posibles de una variable, al contrario del Excel, que se realiza una a la vez.

Ya calculado los estadísticos descriptivos, se muestran como siguiente,

Estadísticos descriptivos de la variable “tiempo de espera en “Tiempo de espera en caja de banco”.

N	Válidos	50
	Perdidos	0
Media		18.00
Mediana		18.00
Moda		15(a)

Desv. típ.		3.625
Varianza		13.143
Asimetría		.019
Error típ. de asimetría		.337
Curtosis		-.455
Error típ. de curtosis		.662
Rango		16
Mínimo		10
Máximo		26
Suma		900
Percentiles	25	15.00
	50	18.00
	75	21.00

a Existen varias modas. Se mostrará el menor de los valores.

Del análisis anterior se puede apreciar que la media del tiempo de espera en caja por un cliente es de 18 minutos y una desviación típica de 3,62 minutos, lo que quiere decir que el intervalo de espera de los clientes en la Caja de un banco, oscila entre [14,38; 21,62].

Se puede observar la coincidencia con el resultado calculado en el ejemplo 10, utilizando la fórmula para el cálculo de la media y con el cálculo realizado en Excel en la figura 111.

Es importante resaltar que utilizando la variante que se entienda más cómoda, los resultados son los mismos al igual que las interpretaciones.

En el ejemplo 14 se tomaron las 10 primeras observaciones de la variable “Tiempo de espera en caja de banco”, si se realiza el cálculo a través del SPSS, obteniéndose:

Cálculo de los estadísticos descriptivos de la variable “Tiempo de espera en caja de banco” utilizando los 10 primeros casos.

N	Válidos	10
	Perdidos	0
Media		18.30
Mediana		18.50

Moda		19
Desv. típ.		4.523
Varianza		20.456
Asimetría		.323
Error típ. de asimetría		.687
Curtosis		-.617
Error típ. de curtosis		1.334
Rango		14
Mínimo		12
Máximo		26
Suma		183
Percentiles	25	14.50
	50	18.50
	75	21.75

Al observarse los cálculos realizados en el ejemplo 14, puede corroborarse que los estadísticos descriptivos coinciden con los calculados con el uso del SPSS.

El SPSS ofrece múltiples posibilidades que ayudan a la riqueza de la investigación, como pueden ser, recodificación de una variable, cruzamiento de dos o más variables, agrupar por grupos de variables, entre otras.

¿Cómo recodificar una variable en SPSS?

Se debe seguir la lógica que este material ha seguido durante todo su recorrido, es decir, desarrollar los saberes desde ejemplos prototipos tratado con Excel.

Se quiere recodificar la variable “Tiempo de espera en caja de banco” en aquellas personas que esperan:

1. Menos de 15 minutos
2. Entre 16 y 20 minutos
3. Entre 21 y 24 minutos
4. 25 y más minutos

Como se aprecia la variable de razón se ha convertido en una variable ordinal de cuatro opciones.

Pasos a seguir en el SPSS para recodificar la variable “Tiempo de espera en caja de banco”

1. El recorrido en SPSS, es el siguiente: Panel de Transformar/ recodificar en distinta variable.

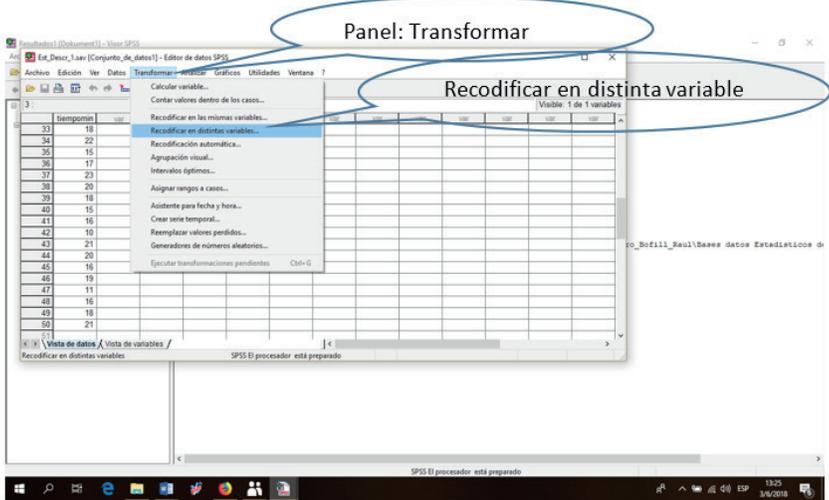


Figura 112. Ventana que genera el SPSS en el panel transformar para recodificar una variable.

2. Se muestra una ventana como sigue:

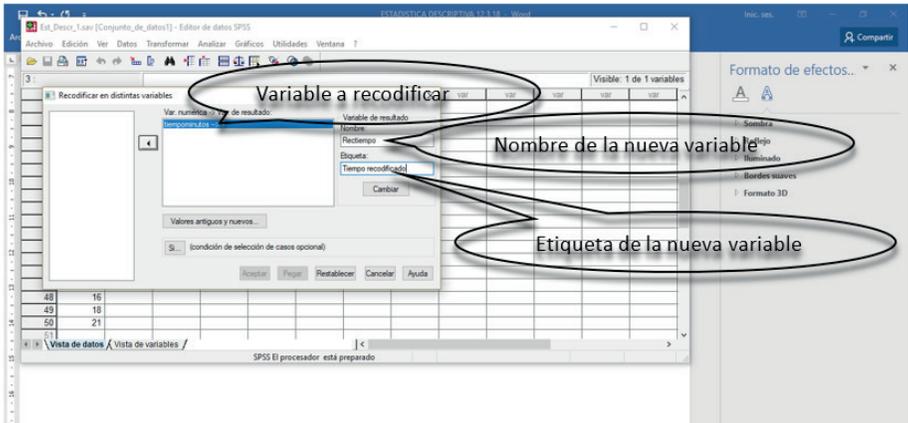


Figura 113. Venta de SPSS para seleccionar la variable que debe ser recodificada.

Como se observa en la figura 113 lo primero que se hace es la variable que se va a recodificar pasarla al cuadro de diálogo “Var. de resultado”, se le asigna un nuevo nombre a la variable resultante, pues en SPSS no puede existir dos variables con el mismo nombre, se agrega una etiqueta para diferenciarla de la primaria u original y se oprime el botón “Valores antiguos”, se prosigue según la figura 114 del tercer paso:

Paso 3:

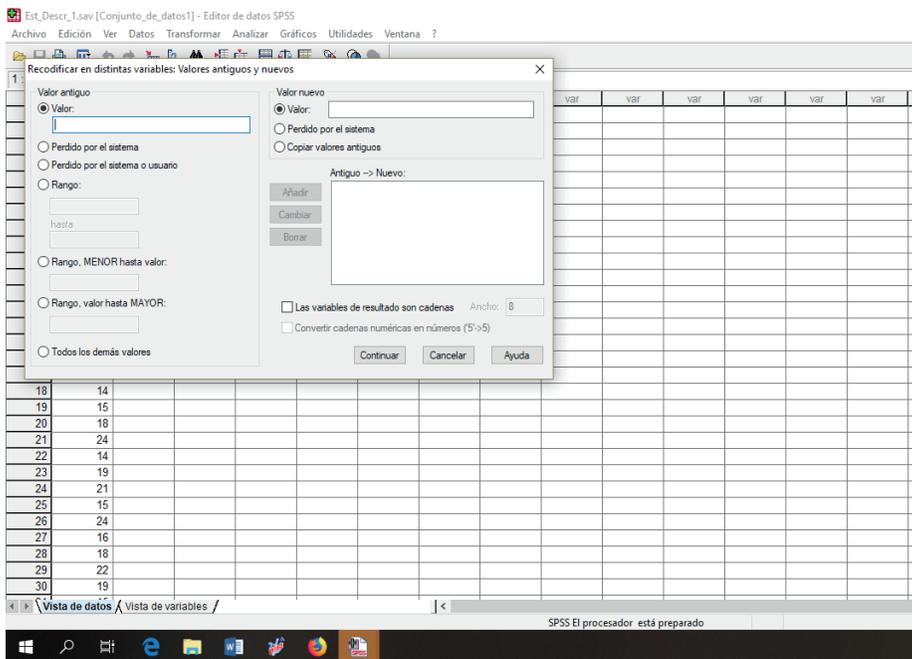


Figura 114. Venta de SPSS para seleccionar “Valores antiguos”.

En la figura 114 se muestra una ventana que a la izquierda tiene “Valores antiguos” y a la derecha “Valor nuevo”, se explicará a continuación todas las posibles opciones:

Valor: es seleccionar un solo valor de la variable antigua para asignarle un nuevo valor en “Valores nuevo”

Perdido por el sistema y Perdido por el sistema o usuario, se utiliza en raras ocasiones por ello no vamos a tratarlo en este apartado.

Rango: al seleccionar esta opción se activa “hasta”, usted debe seleccionar

la amplitud del intervalo y en “Valor” de “Valores nuevos” asignar el valor del rango seleccionado.

Rango, MENOR hasta valor: significa que desde el menor número que tenga el conjunto de datos hasta, un valor determinado, se forma un intervalo y en “Valor” de “Valores nuevos” asignar el valor del rango seleccionado.

3.4. Problemas propuestos para solución

PP 11. En el siguiente conjunto de datos, se proporcionan los pesos (en libras) de borregos nacidos en cierto intervalo de tiempo:

9	7	6	10	8	5	9	6	3	7	5	5
4	8	4	6	8	6	7	7	7	8	10	7
6	4	7	6	9	7	4	7	6	8	8	4
8	9	11	8	7	10	8	5	7	7	9	6

a) Construya una distribución de frecuencia de estos pesos.

b) Encuentre las frecuencias relativas.

c) Determine las frecuencias acumuladas.

d) Calcule las frecuencias relativas acumuladas.

e) Construya un Histograma.

f) Construya un Diagrama de Caja

PP 12. Tomando como base los datos que se dan a continuación referidos a un estudio sobre las características o variables asociadas al personal de una Empresa y utilizando las facilidades del EXCEL, dé solución a las tareas que se plantean.

No. Orden	Sexo	Grado escolar	Categoría Ocupacional	Salario/día	Experiencia Previa (Meses)
1	Masculino	15	Directivo	\$57,000	144
2	Masculino	16	Administrativo	\$40,200	36
3	Femenino	12	Administrativo	\$21,450	381
4	Femenino	8	Administrativo	\$21,900	190
5	Masculino	15	Seguridad	\$45,000	138

6	Masculino	15	Administrativo	\$32,100	67
7	Masculino	15	Administrativo	\$36,000	114
8	Femenino	12	Administrativo	\$21,900	2
9	Femenino	15	Seguridad	\$27,900	115
10	Femenino	12	Administrativo	\$24,000	244
11	Femenino	16	Administrativo	\$30,300	143
12	Masculino	8	Seguridad	\$28,350	26
13	Masculino	15	Administrativo	\$27,750	34
14	Femenino	15	Administrativo	\$35,100	137
15	Masculino	12	Seguridad	\$27,300	66
16	Masculino	12	Administrativo	\$40,800	24
17	Masculino	15	Administrativo	\$46,000	48
18	Masculino	16	Directivo	\$103,750	70
19	Masculino	12	Administrativo	\$42,300	103
20	Femenino	12	Administrativo	\$26,250	48
21	Femenino	16	Administrativo	\$38,850	17
22	Masculino	12	Administrativo	\$21,750	315
23	Femenino	15	Seguridad	\$24,000	75
24	Femenino	12	Administrativo	\$16,950	124
25	Femenino	15	Administrativo	\$21,150	171

Calcular:

- a) Estadígrafos de Posición: Cuartiles
- b) Estadígrafos de Tendencia central: Media Aritmética, Mediana y Moda.
- c) Estadígrafos de Variación: Varianza Muestral y Desviación estándar
- d) Asimetría: Simétrica, Asimétrica positiva ó Asimétrica negativa.
- e) Curtosis: Mesocúrtica, Leptocúrticaó, Platicúrtica.
- f) Tablas de distribución de frecuencias para las Variables: Sexo, Grado escolar y Categoría Ocupacional.
- g) Representación Gráfica: para todas las variables de acuerdo al tipo que le corresponda.

PP 13. En un estudio realizado a las MIPYMES de la Provincia de El Oro, Ecuador, se ha recogido la siguiente información:

Gasto mensual	Ganancia Mensual	Ubicación de la empresa: 1- Rural 2- Urbana	Work Engagement de la MIPYMES 1- Muy Insatisfecho 2- Bastante Insatisfecho 3- Algo insatisfecho 4- Indiferente 5- Algo satisfecho 6- Bastante satisfecho 7- Muy satisfecho
1022	601	1	2
3000	1400	2	4
2133	1212	2	3
1266	812	1	2
1700	801	2	3
2745	1320	2	2
960	520	1	3
2400	1030	1	2
1756	686	1	4
2788	1233	2	5
1444	952	1	5
2345	1000	2	2
1920	965	1	2
989	578	1	3
2800	1023	2	2
2145	996	4	2
1879	1110	2	3
1744	956	1	3
1200	825	1	2
2500	1233	2	2
1800	950	1	3
2455	1002	2	2
2710	1010	2	6
2900	1200	2	1
2800	1400	2	6
1478	788	1	4

2500	1100	2	7
2000	850	1	7
2645	966	2	4

Calcular, según las medidas de las variables:

- a) Estadígrafos de Posición: Cuartiles
- b) Estadígrafos de Tendencia central: Media Aritmética, Mediana y Moda.
- c) Estadígrafos de Variación: Varianza Muestral y Desviación estándar
- d) Asimetría: Simétrica, Asimétrica positiva ó Asimétrica negativa.
- e) Curtosis: Mesocúrtica, Leptocúrticaó, Platicúrtica.
- f) Tablas de distribución de frecuencias para las cuatro variables
- g) Representación Gráfica: para todas las variables de acuerdo al tipo que le corresponda.
- h) Diga cuál de las categorías 1- Rular o 2- Urbana tiene mayor “Gasto mensual”
- i) Diga cuál de las categorías del Work Engagement de la MIPYMES tiene mayor “ganancia mensual”
- j) Recodifique la variable “Gasto mensual” en tres rangos y calcule su gráfico de cajas o bigotes.

Referencias bibliográficas

- Aragón L, (2016). Estadística en el área de las ciencias sociales y administrativas. Santiago de Chile: Alfaomega.
- Christensen, H. (2012). Estadística Paso a Paso. México: Trillas.
- Coscolluela, A., Turbany, J., & Fornieles, A. (2002). Estadística Descriptiva. Barcelona: UOC Papers. España. Recuperado en <http://www.digitaliapublishing.com/a/6172/estad-stica-descriptiva>
- Ecuador. Instituto Nacional de Estadísticas y Censo. (2012). Censo Nacional Económico. Directorio de Empresas. Quito: INEC.
- García, A., & Simón de Blas, C. (2007). Manual de Estadística. Madrid: Dykinson.
- González, L. (2013). Estadística descriptiva y probabilidad. Bogotá: Jorge Tadeo Lozano.
- Hanke, J. (1997) Estadística para negocios. Madrid: McGraw-Hill/Interamericana.
- Ivars Escortell, A., et al. (2005). Estadística Descriptiva y nociones de Probabilidad. Valencia: Thomson Editores.
- Levine, D., & Berenson, M. (2006). Estadística para Administración. México: Pearson Educación.
- López, R., et al. (2016). Expertos y prospectiva en la investigación pedagógica. Cienfuegos: Universo Sur.
- López, R., et al. (2017). Análisis exploratorio de datos con SPSS. Cienfuegos: Universo Sur.
- Rico, A (2016). Manual avanzado Microsoft Excel. México: Pearson Educación.
- Santoyo, F. (2001). Probabilidad y Estadística para la Gestión Empresarial. Morelia: Universidad Michoacana de San Nicolás de Hidalgo.
- Spiegel, M., & Stephens, L. (2009). Estadística. México: Mc Graw Hill
- Thomas, S. (2014). Basic statistics. Recuperado de <http://ebookcentral.proquest.com>
- Valdivieso, M. (2006). Estadística Descriptiva: Apoyo al estudio independiente. Bogotá: Universidad Pedagógica y Tecnológica de Colombia.

Índice

Prólogo.....	7
Introducción	9
Capítulo I. Conceptos fundamentales en estadística	12
1.1. Diferentes tipos de variables estadísticas	12
1.2. Pasos en un estudio estadístico	13
1.3. Resumir los datos de un estudio estadístico. Distribución de frecuencias	14
1.4. Aplicación de la informática al cálculo de frecuencias	24
1.5. Problemas resueltos de construcción de tablas de distribución frecuencial	30
1.6. Representación gráfica de las distribuciones de frecuencias	59
1.7. Problemas resueltos de construcción de gráficos	75
1.8. Problemas propuestos de distribución de frecuencias y su representación gráfica	88
Capítulo II. Medidas descriptivas	93
2.1. Estadígrafos de tendencia central	93
2.1.1. Media aritmética o Promedio	93
2.1.2. Mediana	98
2.1.3. Trimedia	104
2.1.4. Moda	107
2.2. Estadígrafos de Dispersión	108

2.2.1. Amplitud o Rango	108
2.2.2. Varianza Muestral	108
2.2.3. Desviación Típica o Estándar Muestral	109
2.2.4. Coeficiente de Variación	110
2.3. Estadígrafos de Posición	113
2.3.1. Rango Intercuartil	117
2.3.2. Estadígrafos de forma o distribución	117
Capítulo III. Análisis exploratorio de los datos usando los cuartiles. El diagrama de caja	122
3.1. El Diagrama de Caja en los análisis exploratorios	122
3.2. Problemas resueltos sobre medidas descriptivas	128
3.3. Valoración de los estadísticos descriptivos con el uso del SPSS	153
3.4. Problemas propuestos para solución	161
Referencias bibliográficas	165

La estadística es una herramienta que se utiliza en todas las ciencias, tanto en las llamadas duras como en las blandas, pues recoger, organizar, procesar y analizar la información, propia de cualquier rama del saber. La finalidad de este libro es brindar, a estudiantes y profesionales, los saberes de la estadística básica como herramienta de trabajo útil para su formación y desempeño. Se han utilizados métodos de análisis y síntesis, así como el histórico lógico, ambos del nivel teórico para poder hacer un recorrido en el estudio de esta disciplina y llegar a sintetizar los contenidos robustos de la misma. Se presentan conceptos y definiciones necesarios para la comprensión de la obra, ejercicios, resueltos y propuestos, asociados al entorno orense los cuales tienen sus bases o sustentos en el desarrollo local territorial y donde un número importante de ellos son resultado de la práctica social de esta Provincia. Se aprecia en esta obra como la solución de estos problemas son con la ayuda del Microsoft Excel y el Paquete estadístico para las Ciencias Sociales (SPSS) lo cual ayuda a dedicar mayor tiempo a la interpretación de los resultados, que a los cálculos aritméticos, que solo provocan al público para el cual está dirigido el libro, dispersión de la esencia y lo fundamental es potenciar, en última instancia, el cómo poder interpretar los datos asociados a problemas concretos de la realidad social en que crecen y desarrollan.

EDITORIAL



UNIVERSO
S U R



FUNDACIÓN
METROPOLITANA
Fomentando la Educación Superior

ISBN: 978-959-257-549-3



9 789592 575493