

# Análisis Exploratorio

**Triola, (2004). Definición (Análisis exploratorio de datos):** proceso para utilizar herramientas estadísticas (como gráficas, medidas de tendencia central y medidas de variación), con la finalidad de investigar conjuntos de datos para comprender sus características importantes.

Al estar analizando los datos para poder describirlos, cabe la posibilidad de que se presenten datos distantes o datos outlier (extremos o atípicos); es decir, son datos que se encuentran muy alejados del común de los datos, esto es un dato distante que se presenta muy extremo en relación a los demás datos, hay que considerar que los datos distantes afectan en gran medida al valor de la media, o de la desviación estándar; así mismo afectan a los histogramas pues se hace un cambio abrupto en el valor de los datos.

Retomando el ejemplo de los días del cuadro ulceroso, este es un caso donde aparece un dato distante pues observemos que los datos son 2, 2, 3, 3, 3, 3, 4, 24 y el dato 24 está muy por encima de los datos restantes, esto afectará el promedio pues se tiene que  $\bar{x} = \frac{44}{8} = 5.5$  y si no consideráramos al dato 24, el promedio de días sería  $\bar{x} = \frac{20}{7} = 2.85$  días, por lo que se observa que el número 24 es un dato distante que afecta el análisis exploratorio.

Un procedimiento sencillo para observar los datos distantes es ordenar los datos en una tabla y observar los datos mínimos y máximos y si sobre sale mucho un valor, quiere decir que es un valor distante pues se alejan mucho de los demás datos.

Según John W. Tukey (E.D.A.: Exploratory data analysis) el análisis exploratorio es, básicamente, el tratamiento estadístico al que se someten las muestras recogidas durante un proceso de investigación en cualquier campo científico. Es conveniente mencionar que para mayor rapidez y precisión, todo el proceso suele realizarse por medios informáticos (si se tiene el acceso a estos). Por ejemplo, el software que realiza cálculos estadísticos importantes es Excel y tiene la bondad de que en la mayoría de las computadoras ya se encuentra instalado de fábrica, o bien la opción gratuita de Microsoft office es el Libre Office, que es una buena opción por la similitud en las funciones pero con la ventaja que es de acceso libre.

# Análisis Exploratorio

Lo benéfico de tener a la mano un paquete estadístico y saber manipularlo es que ahorra en gran medida los cálculos que a mano a veces se vuelven tediosos, de igual manera las gráficas quedan mejor presentadas hechas en un paquete estadístico por sencillo que éste sea, que a mano y tus tareas o reportes no demeritarán, al contrario se destacará la habilidad no solo estadística sino computacional.

A continuación se muestra un ejemplo de una base de datos en Excel:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL		
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35				
2	Encuestado	Letivación	gureña	edad	toiva	en	usate	tercipes	nuda	pasuip	cial	vidyuda	restidear	maza	ramon	pm	apmer	maon	s	pp	ln	dear																	Edad	Género
3	1	6	7	8	9	6	3	7	6	5	9	4	8	2	9	1	0	10	8	7	9	8	2	2	1	9	9	7	8	6	5	10	9	8	5	4	16	1		
4	2	7	6	8	7	8	1	2	9	9	8	8	9	6	10	5	3	9	8	7	9	7	8	5	6	10	10	9	8	8	9	10	9	8	9	7	16	2		
5	3	10	9	10	7	5	9	5	8	9	10	8	10	1	10	9	6	10	10	5	10	5	5	6	8	9	6	7	6	5	10	10	9	8	7	6	16	1		
6	4	9	10	10	9	9	9	8	9	9	9	9	8	10	8	8	10	10	9	10	10	9	10	10	9	10	9	10	10	10	10	9	10	9	9	9	9	16	2	
7	5	7	9	5	9	8	10	9	9	8	9	7	10	0	9	0	3	7	6	5	9	0	8	2	2	7	4	0	8	8	10	7	8	8	7	6	16	1		
8	6	9	10	8	10	8	8	9	9	9	10	8	10	7	10	7	6	8	9	8	7	9	8	9	7	8	10	10	9	10	10	10	9	10	8	8	16	2		
9	7	10	7	6	8	5	4	0	2	1	3	10	10	9	8	7	9	3	3	7	2	1	9	10	9	8	4	3	9	0	10	0	1	10	8	7	15	2		
10	8	10	8	7	9	6	7	8	9	10	8	6	10	7	10	6	6	10	10	7	8	8	1	9	8	10	9	8	7	8	7	9	10	2	7	6	15	2		
11	9	8	8	7	9	10	10	7	8	9	7	9	6	10	7	6	10	8	7	10	8	7	7	6	9	8	10	9	8	10	7	7	9	7	9	15	2			
12	10	10	9	10	7	10	10	10	10	10	7	5	5	7	10	0	0	10	7	0	10	8	6	0	0	5	0	1	6	8	10	0	6	4	2	3	15	2		
13	11	9	7	6	5	9	8	10	9	9	8	8	10	7	10	8	7	6	7	9	10	8	8	7	6	8	7	10	10	10	9	7	9	10	10	6	15	2		
14	12	9	9	10	9	8	9	7	2	10	0	0	1	0	1	0	0	8	2	0	10	2	4	3	6	4	5	6	7	8	4	4	4	4	4	4	15	2		
15	13	8	7	3	2	8	6	1	1	6	8	6	7	6	8	6	3	8	8	3	8	4	3	3	3	3	8	7	4	3	8	8	7	3	4	5	16	2		
16	14	8	9	6	9	10	5	7	10	5	9	9	10	8	9	8	9	6	5	10	8	9	10	9	10	9	7	8	7	6	9	9	10	9	10	8	16	2		
17	15	8	10	7	1	5	7	9	10	7	5	9	10	7	10	6	9	8	7	8	9	7	5	5	8	9	6	9	10	10	5	9	9	5	7	5	16	2		
18	16	9	10	10	7	8	6	6	10	10	9	8	10	6	10	6	10	10	7	10	10	7	6	7	6	7	8	9	6	5	10	10	9	10	9	8	16	2		
19	17	9	10	7	10	9	8	8	7	6	8	9	10	10	9	8	7	8	9	10	10	9	5	10	9	5	10	7	8	8	8	10	8	9	10	8	15	2		
20	18	10	7	8	10	8	7	9	10	7	8	9	8	9	7	8	9	9	9	8	7	7	8	6	8	5	7	7	8	9	9	9	9	10	10	9	16	2		
21	19	9	8	9	7	8	5	6	3	3	9	8	8	10	5	5	10	9	10	10	8	2	2	2	7	8	9	7	6	8	10	9	8	7	8	16	2			
22	20	8	7	7	6	8	9	7	9	6	5	8	9	7	7	9	7	6	7	9	10	6	9	3	4	5	7	8	9	6	5	3	4	7	8	7	15	2		
23	21	7	8	6	7	7	7	7	7	8	6	6	6	7	7	8	6	7	10	7	8	9	7	8	6	1	8	8	8	8	9	8	9	8	9	8	15	2		
24	22	9	10	10	8	9	9	10	9	10	10	8	9	9	10	10	9	10	10	8	10	9	8	7	8	10	9	8	9	9	9	10	10	9	8	5	15	2		
25	23	10	7	8	9	6	7	8	9	10	5	8	9	7	6	8	6	5	8	10	9	7	7	10	8	5	6	7	8	9	10	9	10	8	7	10	17	2		
26	24	10	9	7	8	5	6	3	4	1	2	7	9	10	5	8	6	4	2	3	1	4	10	10	8	5	9	0	10	9	10	9	10	7	6	10	15	2		
27	25	8	10	8	10	9	5	10	5	10	4	10	10	10	10	10	8	5	8	9	10	10	10	10	10	10	7	10	10	10	10	9	10	10	9	8	15	2		

# Análisis Exploratorio

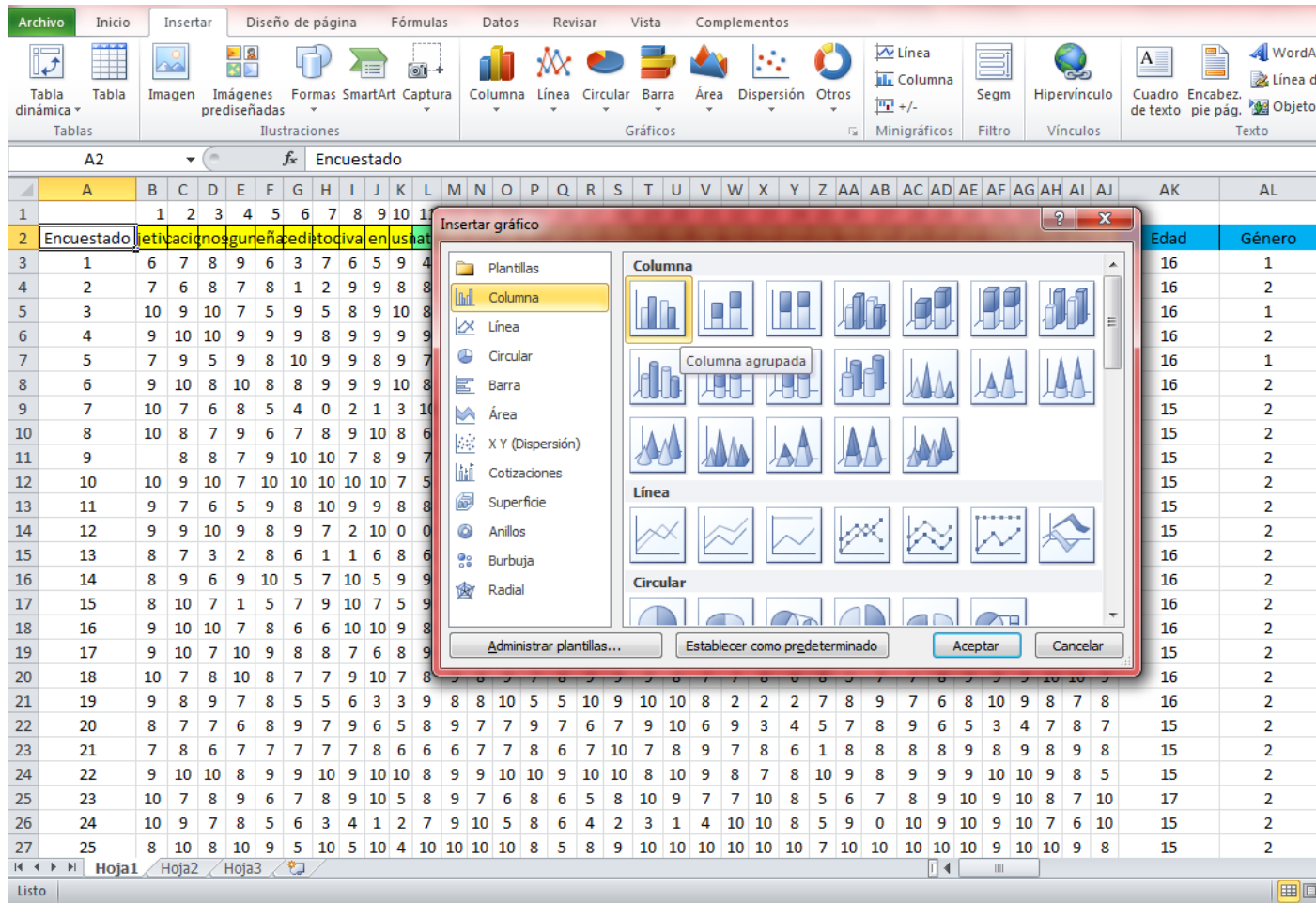
Y a partir de las fórmulas estadísticas, se pueden obtener diversos indicadores para poder hacer un buen análisis:

The screenshot shows the Microsoft Excel interface with the 'Formulas' ribbon selected. The 'Biblioteca de funciones' (Function Library) is open, showing the 'Estadísticas' (Statistical) category. The background spreadsheet contains a table of survey data with columns for 'Edad' (Age) and 'Género' (Gender).

Encuestado	Edad	Género
1	16	1
2	16	2
3	16	1
4	16	2
5	16	1
6	16	2
7	16	1
8	16	2
9	15	2
10	15	2
11	15	2
12	15	2
13	15	2
14	15	2
15	16	2
16	16	2
17	16	2
18	16	2
19	15	2
20	16	2
21	16	2
22	16	2
23	15	2
24	15	2
25	17	2
26	15	2
27	15	2

# Análisis Exploratorio

Así mismo se pueden realizar diversos tipos de gráficas: líneas, columnas, circulares, entre otras, para tener una representación visual de los datos:



**Referencia:**

Rivera, E (2014) Análisis Exploratorio [elsarivera@uadec.edu.mx](mailto:elsarivera@uadec.edu.mx)  
Triola, M (2004) Probabilidad y Estadística Novena Edición, Pearson Educación, México.