



Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Sociología

Documento de Cátedra 25: Introducción a la estadística descriptiva: principales conceptos

Autores: Gelfman, Julia; Martínez Mendoza, Ramiro; Rosa, Paula y Tignino, María Victoria (2007)

Cátedra de Metodología y Técnicas de la Investigación Social, Profesora Titular: Ruth Sautu, Carrera de Sociología, Facultad de Ciencias Sociales, Universidad de Buenos Aires (UBA)

Este Documento de Cátedra forma parte de una serie que tiene como propósito contribuir a la formación de los alumnos de la Carrera de Sociología de la Universidad de Buenos Aires en cuestiones vinculadas al diseño y realización de investigaciones científicas en el campo de las ciencias sociales. Su contenido complementa los textos de metodología de lectura obligatoria y optativa incluidos en cada uno de los tres cursos que se dictan en la Carrera.

La Cátedra solicita a los usuarios de estos Documentos que citen a sus autores indicando las referencias completas, es decir: autores, fecha, título y número del documento (traducción, resumen, elaboración propia, etc.). En los casos en que el usuario utilice sólo parte del documento, haciendo referencia a algunos de los autores/obras originales allí incluidos, consignar que fue tomado de nuestro Documento de Cátedra. Por ejemplo:

E.O Wrigth (1985), *Classes*, London: Verso, citado en Documento de Cátedra II.1., Plotno, G., Lederman, F. & Krause, M. (2007) "Escalas Ocupacionales".

I. Introducción

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

Las **variables** pueden ser de dos tipos:

- *Variables cualitativas*: no se pueden medir numéricamente (por ejemplo: nacionalidad, color de la piel, sexo).
- *Variables cuantitativas*: tienen valor numérico (edad, precio de un producto, ingresos anuales).

A su vez cada uno de estos tipos de variables pueden distinguirse entre:

Variables cualitativas:

- *Nominales*: refieren a diferencias en cualidades o denominaciones, que no implican un ordenamiento de ningún tipo entre las categorías de la misma variable.
- *Ordinales*: refieren a variables con sistemas de categorías que involucran un ordenamiento, por ejemplo de mayor a menor.

Variables cuantitativas:

- *Intervalares*: poseen un sistema de categorías numérico, que permite identificar diferencias en cantidad, monto, etc. Además, en sus puntuaciones el cero **no implica** la ausencia de la cualidad o atributo que se está midiendo.
- *De cociente o razón*: posee las mismas características que las variables de intervalo, sin embargo en sus puntuaciones el cero **implica** la ausencia de la cualidad o atributo que se está midiendo.

Cuando se estudia el comportamiento de una variable hay que distinguir los siguientes conceptos:

- *Individuo*: cualquier elemento que aporte información sobre el fenómeno que se estudia. Así, si estudiamos la altura de los niños de una clase, cada alumno es un individuo; si estudiamos el precio de la vivienda, cada vivienda es un individuo.
- *Población*: conjunto de todos los individuos (personas, objetos, animales, etcétera) que porten información sobre el fenómeno que se estudia. Por ejemplo, si estudiamos el precio de la vivienda en una ciudad, la población será el total de las viviendas de dicha ciudad.
- *Unidad de análisis*: Son es el *qué* o el *quién* se quiere investigar. Es decir, aquellas unidades u objetos que examinamos con el fin de crear descripciones. Generalmente en las ciencias sociales las unidades de análisis son personas.

(Babbie; 1996). Existen distintos tipos de unidades de análisis. Entre los más habituales para las ciencias sociales podemos encontrar:

Individuos: Las personas son las unidades de análisis más típicas. Cualquier variedad de individuos se puede considerar unidad de análisis. Los individuos pueden ser caracterizados según su pertenencia a grupos sociales.

Grupos: Los grupos sociales pueden ser en sí mismos unidades de análisis para la investigación en ciencias sociales. Las familias podrían ser las unidades de análisis de un estudio. También los matrimonios, las ciudades o las regiones geográficas. Cada uno de estos términos implica algún grupo de población.

Organizaciones: Las organizaciones sociales formales pueden ser unidades de análisis. Las iglesias, los supermercados o las universidades forman parte de este tipo de unidad de análisis.

Artefactos sociales: Los artefactos sociales o los productos de la vida o la conducta social también pueden ser unidades de análisis. Pueden ser los libros, las pinturas, las canciones o los juegos. Cada uno de estos objetos implica una población de todos ellos. Otro tipo de artefactos sociales son las interacciones sociales, como ejemplo podemos nombrar las bodas, los accidentes de tránsito, las peleas o los secuestros aéreos.

- **Muestra:** subconjunto que seleccionamos de la población. Así, si se estudia el precio de la vivienda de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (sería una labor muy compleja), sino que se suele seleccionar un subgrupo (muestra) que se entienda que es suficientemente representativo.

II. Distribución de frecuencia

La **distribución de frecuencia** es la representación estructurada, en forma de tabla, de toda la información que se ha recogido sobre la variable que se estudia.

Veamos un ejemplo: Medimos la edad (en años cumplidos a la fecha del relevamiento) de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005 y obtenemos los siguientes resultados:

Tabla 1.1: Distribución de frecuencias de la variable edad (en años cumplidos a la fecha del relevamiento) de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Joven Investigador	Edad
x-unidad de análisis	x
JI Nº 1	29
JI Nº 2	25
JI Nº 3	34
JI Nº 4	28
JI Nº 5	25
JI Nº 6	26
JI Nº 7	23
JI Nº 8	31
JI Nº 9	26
JI Nº 10	24
JI Nº 11	33
JI Nº 12	33
JI Nº 13	37
JI Nº 14	26
JI Nº 15	25
JI Nº 16	37
JI Nº 17	38
JI Nº 18	27
JI Nº 19	24
JI Nº 20	36
JI Nº 21	23
JI Nº 22	36
JI Nº 23	29
JI Nº 24	26
JI Nº 25	27
JI Nº 26	27
JI Nº 27	40
JI Nº 28	35
JI Nº 29	29
JI Nº 30	29
JI Nº 31	25
JI Nº 32	27
JI Nº 33	29
JI Nº 34	31
JI Nº 35	29
JI Nº 36	30
JI Nº 37	30
JI Nº 38	25
JI Nº 39	27
JI Nº 40	26
JI Nº 41	28
JI Nº 42	24
JI Nº 43	26
JI Nº 44	28
JI Nº 45	24
JI Nº 46	32
JI Nº 47	30
JI Nº 48	28
JI Nº 49	40
JI Nº 50	31
JI Nº 51	25
JI Nº 52	27
JI Nº 53	24
JI Nº 54	29
JI Nº 55	26
JI Nº 56	30
JI Nº 57	28
JI Nº 58	29
JI Nº 59	28
JI Nº 60	32
JI Nº 61	34
JI Nº 62	31
JI Nº 63	30
JI Nº 64	30
JI Nº 65	32
JI Nº 66	32
JI Nº 67	30
JI Nº 68	24
JI Nº 69	28
JI Nº 70	29
JI Nº 71	28
JI Nº 72	28
JI Nº 73	26
JI Nº 74	30
JI Nº 75	28
JI Nº 76	25
JI Nº 77	28
JI Nº 78	31
JI Nº 79	27
JI Nº 80	28
JI Nº 81	33
JI Nº 82	30
JI Nº 83	28
JI Nº 84	25
JI Nº 85	29
JI Nº 86	27
JI Nº 87	26
JI Nº 88	27
JI Nº 89	25
JI Nº 90	35
JI Nº 91	23
JI Nº 92	28
JI Nº 93	24
JI Nº 94	28
JI Nº 95	25
JI Nº 96	31
JI Nº 97	29
JI Nº 98	27
JI Nº 99	25
JI Nº 100	27
JI Nº 101	32
JI Nº 102	27
JI Nº 103	26
JI Nº 104	30
JI Nº 105	40
JI Nº 106	31
JI Nº 107	26
JI Nº 108	35
JI Nº 109	29
JI Nº 110	36
JI Nº 111	28
JI Nº 112	30
JI Nº 113	27
JI Nº 114	35
JI Nº 115	27
JI Nº 116	26
JI Nº 117	29
JI Nº 118	29
JI Nº 119	31
JI Nº 120	28
JI Nº 121	28
JI Nº 122	34
JI Nº 123	28
JI Nº 124	33
JI Nº 125	34
JI Nº 126	38
JI Nº 127	35
JI Nº 128	36
JI Nº 129	39
JI Nº 130	36
JI Nº 131	34
JI Nº 132	39
JI Nº 133	40
JI Nº 134	31
JI Nº 135	37
JI Nº 136	29
JI Nº 137	32
JI Nº 138	32
JI Nº 139	37
JI Nº 140	30
JI Nº 141	32
JI Nº 142	32
JI Nº 143	36
JI Nº 144	35
JI Nº 145	36
JI Nº 146	37
JI Nº 147	23
JI Nº 148	30
JI Nº 149	38
JI Nº 150	32
JI Nº 151	40
JI Nº 152	35
JI Nº 153	33
JI Nº 154	33

Fuente: Elaboración propia

Si presentamos esta información estructurada obtendríamos la siguiente **tabla de frecuencia**:

Tabla 1.2: Distribución de frecuencias de la variable edad (en años cumplidos a la fecha del relevamiento) de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Variable (Valor)	Frecuencias absolutas		Frecuencias porcentuales	
	Simple	Acumulada	Simple	Acumulada
23	4	4	2,6%	2,6%
24	7	11	4,5%	7,1%
25	11	22	7,1%	14,3%
26	12	34	7,8%	22,1%
27	14	48	9,1%	31,2%
28	19	67	12,3%	43,5%
29	15	82	9,7%	53,2%
30	13	95	8,4%	61,7%
31	9	104	5,8%	67,5%
32	10	114	6,5%	74,0%
33	6	120	3,9%	77,9%
34	5	125	3,2%	81,2%
35	7	132	4,5%	85,7%
36	7	139	4,5%	90,3%
37	5	144	3,2%	93,5%
38	3	147	1,9%	95,5%
39	2	149	1,3%	96,8%
40	5	154	3,2%	100,0%
Total	154		100,0%	

Fuente: Elaboración propia

Si los valores que toma la variable son muy diversos y cada uno de ellos se repite muy pocas veces, entonces conviene agruparlos por intervalos, ya que de otra manera obtendríamos una tabla de frecuencia muy extensa que aportaría muy poco valor a efectos de síntesis. (Tal como se verá a continuación).

III. Distribuciones de frecuencia agrupada

Supongamos que medimos la edad de Jóvenes investigadores de la Universidad de Buenos Aires en el año 2005 y obtenemos los siguientes resultados (en años cumplidos a la fecha del relevamiento):

Tabla 1.3: Distribución de frecuencias de la variable edad (en años cumplidos a la fecha del relevamiento) de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Edad (en años cumplidos a la fecha del relevamiento)	Frecuencia	Porcentaje	Porcentaje acumulado
20 - 25	22	14,3	14,3
26 - 30	73	47,4	61,7
31 - 35	37	24,0	85,7
36 - 40	22	14,3	100,0
Total	154	100,0%	

Fuente: Elaboración propia

Si al presentar esta información en una tabla hubiésemos obtenido como resultado de frecuencias una tabla de 154 filas (una para cada valor), cada una de ellas con una frecuencia absoluta de 1, dicha tabla nos aportaría escasa o nula información para ser analizada. En lugar de ello, se agruparon los datos por intervalos, quedando la información más resumida.

Los procesos de agrupamiento de datos, donde la información es presentada de manera agregada, implican una pérdida de información detallada pero al mismo tiempo dicho proceso provee al investigador de información más manejable posibilitando su análisis.

El número de tramos en los que se agrupa la información es una decisión que debe tomar el investigador. La regla es que mientras más tramos se utilicen menos información se pierde, pero puede que menos representativa e informativa sea la tabla.

III.I Punto Medio

En el caso de los datos agrupados (especialmente en el caso del uso de intervalos) no tenemos el valor concreto de los datos para cada unidad de análisis, por lo tanto usamos el punto medio del intervalo (que funcionaría como una especie de promedio del intervalo). El punto medio es el promedio entre los límites superior e inferior del intervalo.

Cálculo del punto medio:

$$\frac{\text{Límite inferior} + \text{límite superior}}{2} = \text{Punto medio}$$

En nuestro ejemplo (ver tabla 1.3), el valor del punto medio para el **intervalo 20 a 25 años**:

$$\frac{20 + 25}{2} = 22.5$$

IV. Proporción, Porcentaje, Razón y Tasa

Siguiendo a Holguín Quiñónez (1988) podemos decir que el **porcentaje** es la relación que se establece entre una de las partes con respecto al todo o total multiplicado por cien. La utilidad de los porcentajes está dada porque de ellos se puede obtener *comparabilidad* que es posible porque los números absolutos se reducen a una escala que es fácil de multiplicar y dividir.

Un error común en la utilización de porcentajes es manipular los porcentajes como si estos fuesen números absolutos. Cuando los porcentajes fueron obtenidos de bases diferentes no se los debe sumar ni promediar ni combinar. Solo es correcto realizar estas operaciones si provienen de una misma base de datos.

Los porcentajes deben ser calculados en el sentido del factor que se considera como causa (variable independiente). La comparación debe realizarse en el sentido contrario.

Cálculo del porcentaje:

$$\frac{n}{N} * 100 = \%$$

En nuestro ejemplo (ver tabla 1.2), para el valor **23 años**:

$$\frac{4}{154} * 100 = 2.6\%$$

Las **proporciones**, son la relación que se establece entre una parte con respecto al todo. En las proporciones no se multiplica por cien, es por esta razón que la relación se establece con respecto a la unidad. Para Ritchey (2001), la proporción parte de la cantidad total o número de observaciones, expresada en decimal.

Cálculo de la proporción:

$$\frac{n}{N} = \text{proporción}$$

En nuestro ejemplo (ver tabla 1.3), para el intervalo **26 - 30 años**:

$$\frac{73}{154} = 0,47$$

Para Holguín Quiñónez (1988), las **razones** son una relación que se establece entre dos subconjuntos o dos conjuntos. Si las razones se multiplican o dividen por un mismo número no se alteran, lo cual permite expresarlas como números enteros en ocasiones. En demografía se utiliza mucho la razón de hombres sobre mujeres.

Tabla 1.4: Distribución de frecuencias de la variable Sexo de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Sexo	Frecuencia	Porcentaje
Mujer	69	44,8
Varón	85	55,2
Total	154	100,0%

Fuente: Elaboración propia

Cálculo de la razón:

$$\frac{A}{B} = \text{razón}$$

En nuestro ejemplo (ver tabla 1.4), la razón de mujeres sobre varones es de:

$$\frac{69}{85} = 0,81$$

Lo que este valor nos está indicando, es que por cada 10 varones investigadores encontramos 8 mujeres investigadoras.

En el caso de la **tasa**, Ritchey (2001) sostiene que es la frecuencia de ocurrencia de un fenómeno en relación a un número “base” especificado de sujetos de una población. Para Holguín Quiñónez (1988), en una tasa o coeficiente el numerador indica el número de veces que un evento específico ocurre durante un particular período de tiempo y el denominador es el número de veces que el evento está sujeto al riesgo de que ocurra o acontezca. Por lo general las tasas se multiplican por algún número (mil, 10 mil, 100 mil, etc.) para expresarlas por el número de veces que el evento del numerador acontece en cada, por ejemplo 1000 casos. Las usualmente conocidas son las tasas de mortalidad y natalidad.

Cálculo de la tasa:

$$(A / B) * 1000 = \text{tasa}$$

En nuestro ejemplo calcularemos la *tasa de investigadores*. La tasa de investigadores es el cociente entre el número de investigadores en un determinado período de tiempo (en general un año) y el total de la población en nuestro caso expresada cada mil habitantes.

Total de investigadores en el país al 2001¹: **52.243**

Población total del país al 2001²: **36.260.130**

Tasa de investigadores: $(52243 / 36260130) * 1000 = 1,4$

Lo que este valor nos está indicando, es que en el año 2001 la argentina tenía una persona (1,4) dedicada a investigación y desarrollo cada mil habitantes.

V. Medidas de tendencia central

Entre las medidas de tendencia central más comunes podemos encontrar el modo, la *media aritmética*, la *mediana*. Según Ritchey (2001), un estadístico de tendencia central “proporciona una estimación de la puntuación típica, común o normal encontrada en una distribución de puntuaciones en bruto”.

- *Modo*: se aplica a variables de niveles de medición nominal, ordinal, intervalares o de cociente o razón. El modo es el valor de la variable que se repite más veces en una distribución. Es decir, que es la puntuación con mayor frecuencia. Cabe aclarar que el modo es más eficiente cuando se la utiliza en relación con la mediana y la media aritmética.

Tradicionalmente se lo simboliza *Mo*.

¹ Fuente: Ministerio de Educación, Ciencia y Tecnología. Secretaría de Ciencia, Tecnología e Innovación Productiva. Dirección Nacional de Planificación y Evaluación.

² Fuente: INDEC, Censo Nacional de Población, Hogares y Viviendas 2001.

Tabla 1.5: Distribución de frecuencias de la variable Nombre de la Carrera de Grado de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Nombre de la Carrera de Grado	Frecuencia	Porcentaje
Sociología	37	24,0
Comunicación Social	2	1,3
Bioquímica	38	24,7
Farmacia	9	5,8
Química /Ciencias Químicas	9	5,8
Oceanografía	1	0,6
Biología	28	18,2
Ciencias Políticas	6	3,9
Física	9	5,8
Matemática	5	3,2
Geología / Ciencias geológicas	3	1,9
Nutrición	1	0,6
Derecho	3	1,9
Trabajo Social	1	0,6
Composición Coreográfica	1	0,6
Medicina	1	0,6
Total	154	100,0%

Fuente: Elaboración propia

El *Mo* de la distribución es **Bioquímica**, ya que la mayor cantidad de unidades de análisis (38) cursaron esa carrera de grado.

- *Media aritmética* (o promedio): se aplica a las variables intervalares o de cociente o razón. Para obtener la media se realiza una suma de las puntuaciones y se las divide por el total de puntuaciones. De este modo se obtiene el promedio de una distribución.

La media es considerada como el *punto de equilibrio* de las puntuaciones. Sin embargo, se puede hallar influenciada por los valores extremos negativos como positivos de una distribución.

Tradicionalmente se simboliza de este modo: \bar{X}

Cálculo de la *media aritmética*:

$$\frac{\sum n}{N} = \bar{X}$$

Tomando como ejemplo la distribución de frecuencias de la variable Edad descrita en la Tabla 1.1:

$$\frac{\begin{array}{l} (29 + 25 + 34 + 38 + 25 + 26 + 23 + 31 + 26 + 24 + 33 + 33 + 37 + 26 + 25 + 37 + \\ 38 + 27 + 24 + 36 + 23 + 30 + 29 + 26 + 27 + 27 + 40 + 35 + 29 + 29 + 25 + 27 + \\ 29 + 31 + 29 + 30 + 30 + 25 + 27 + 26 + 28 + 24 + 26 + 28 + 24 + 32 + 30 + 28 + \\ 40 + 31 + 25 + 27 + 24 + 29 + 26 + 30 + 28 + 29 + 28 + 32 + 34 + 31 + 30 + 30 + \\ 32 + 32 + 30 + 24 + 28 + 29 + 28 + 27 + 26 + 30 + 28 + 25 + 28 + 31 + 27 + 28 + \\ 33 + 30 + 28 + 25 + 29 + 27 + 26 + 27 + 25 + 35 + 23 + 28 + 24 + 28 + 25 + 31 + \\ 29 + 27 + 25 + 27 + 32 + 27 + 26 + 30 + 40 + 31 + 26 + 35 + 29 + 36 + 28 + 30 + \\ 27 + 35 + 27 + 26 + 29 + 29 + 31 + 28 + 28 + 34 + 28 + 33 + 34 + 38 + 35 + 36 + \\ 39 + 36 + 34 + 39 + 40 + 31 + 37 + 29 + 32 + 32 + 37 + 30 + 32 + 32 + 36 + 35 + \\ 36 + 37 + 23 + 30 + 38 + 32 + 40 + 35 + 33 + 33 \end{array}}{154} = 30.0$$

$$\frac{4.623}{154} = 30.0$$

El promedio de edad de los jóvenes investigadores de la UBA es de 30 años.

Tomando como ejemplo la distribución de frecuencias que está agrupada, como ese el caso de la variable Edad descrita en la Tabla 1.2:

Cálculo de la *media aritmética para datos agrupados*:

$$\frac{\sum(n * f)}{N} = \bar{X}$$

$$\frac{\begin{array}{l} [(23 * 4) + (24 * 7) + (25 * 11) + (26 * 12) + (27 * 14) + (28 * 19) + (29 * 15) + (30 * 13) + (31 * 9) + (32 * 10) + \\ (33 * 6) + (34 * 5) + (35 * 7) + (36 * 7) + (37 * 5) + (38 * 3) + (39 * 2) + (40 * 5) \end{array}}{154} = 30.0$$

$$\frac{4.623}{154} = 30.0$$

Si la distribución de frecuencias está agrupada en intervalos como el caso de la tabla 1.3 debemos considerar los puntos medios de los intervalos para hacer el cálculo:

Cálculo de la *media aritmética para datos agrupados en intervalos*:

$$\frac{\sum(PM * f)}{N} = \bar{X}$$

$$\frac{[(22,5 * 22) + (28 * 73) + (33 * 37) + (38 * 22)]}{154} = 29.8$$

$$\frac{[(22,5 * 22) + (28 * 73) + (33 * 37) + (38 * 22)]}{154} = 29.8$$

$$\frac{[4.596]}{154} = 29.8$$

Podemos observar que el valor de la Media aritmética no es igual cuando los datos están agrupados en intervalos.

- *Mediana*: se aplica a variables ordinales y a variables cuantitativas. La mediana es el valor o categoría de la variable que divide a la distribución ordenada en dos partes iguales, es decir, es la *puntuación de la mitad*.

Para obtener la mediana es necesario ordenar la distribución (de menor a mayor o de mayor a menor) y luego dividir el total de la distribución (n) entre dos. Cabe aclarar que si el orden de la mediana se localiza entre dos puntuaciones es necesario sacar la media de esas dos puntuaciones para obtener la mediana.

Tradicionalmente se la simboliza *Mdn* o *Md*.

Tabla 1.7: Distribución de frecuencias de la variable Años de Graduado de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Joven Investigador		Años de Graduado
Ud. A. N°	formulario N°	x
1	JI N° 2	1
2	JI N° 5	1
3	JI N° 7	1
4	JI N° 9	1
5	JI N° 10	1
6	JI N° 16	1
7	JI N° 18	1
8	JI N° 19	1
9	JI N° 21	1
10	JI N° 25	1
11	JI N° 30	1
12	JI N° 32	1
13	JI N° 38	1
14	JI N° 42	1
15	JI N° 43	1
16	JI N° 51	1
17	JI N° 53	1
18	JI N° 68	1
19	JI N° 70	1
20	JI N° 76	1
21	JI N° 83	1
22	JI N° 87	1
23	JI N° 91	1
24	JI N° 99	1
25	JI N° 107	1
26	JI N° 116	1
27	JI N° 118	1
28	JI N° 6	2
29	JI N° 15	2
30	JI N° 26	2
31	JI N° 31	2
32	JI N° 36	2
33	JI N° 40	2
34	JI N° 41	2
35	JI N° 45	2
36	JI N° 55	2
37	JI N° 72	2
38	JI N° 79	2
39	JI N° 80	2
40	JI N° 84	2
41	JI N° 89	2
42	JI N° 92	2
43	JI N° 93	2
44	JI N° 95	2
45	JI N° 100	2
46	JI N° 119	2
47	JI N° 121	2
48	JI N° 123	2
49	JI N° 14	3
50	JI N° 24	3
51	JI N° 28	3
52	JI N° 33	3
53	JI N° 34	3
54	JI N° 39	3
55	JI N° 56	3
56	JI N° 59	3
57	JI N° 66	3
58	JI N° 71	3
59	JI N° 94	3
60	JI N° 96	3
61	JI N° 98	3
62	JI N° 102	3
63	JI N° 103	3
64	JI N° 3	4
65	JI N° 4	4
66	JI N° 8	4
67	JI N° 12	4
68	JI N° 23	4
69	JI N° 29	4
70	JI N° 35	4
71	JI N° 47	4
72	JI N° 48	4
73	JI N° 52	4
74	JI N° 57	4
75	JI N° 65	4
76	JI N° 73	4
77	JI N° 77	4
78	JI N° 81	4
79	JI N° 86	4
80	JI N° 106	4
81	JI N° 111	4
82	JI N° 113	4
83	JI N° 114	4
84	JI N° 115	4
85	JI N° 44	5
86	JI N° 60	5
87	JI N° 64	5
88	JI N° 67	5
89	JI N° 69	5
90	JI N° 75	5
91	JI N° 82	5
92	JI N° 88	5
93	JI N° 97	5
94	JI N° 109	5
95	JI N° 112	5
96	JI N° 117	5
97	JI N° 120	5
98	JI N° 142	5
99	JI N° 11	6
100	JI N° 13	6
101	JI N° 17	6
102	JI N° 37	6
103	JI N° 58	6
104	JI N° 62	6
105	JI N° 63	6

106	JI N° 78	6
107	JI N° 85	6
108	JI N° 104	6
109	JI N° 134	6
110	JI N° 136	6
111	JI N° 141	6
112	JI N° 147	6
113	JI N° 150	6
114	JI N° 1	7
115	JI N° 54	7
116	JI N° 74	7
117	JI N° 125	7
118	JI N° 154	7
119	JI N° 20	8
120	JI N° 27	8
121	JI N° 46	8
122	JI N° 101	8

123	JI N° 122	8
124	JI N° 138	8
125	JI N° 140	8
126	JI N° 143	8
127	JI N° 148	8
128	JI N° 152	8
129	JI N° 153	8
130	JI N° 50	9
131	JI N° 108	9
132	JI N° 124	9
133	JI N° 137	9
134	JI N° 61	10
135	JI N° 131	10
136	JI N° 22	11
137	JI N° 90	11
138	JI N° 126	11
139	JI N° 128	11

140	JI N° 145	11
141	JI N° 110	12
142	JI N° 127	13
143	JI N° 135	13
144	JI N° 144	13
145	JI N° 146	13
146	JI N° 149	13
147	JI N° 130	14
148	JI N° 139	14
149	JI N° 132	15
150	JI N° 151	15
151	JI N° 49	16
152	JI N° 129	16
153	JI N° 105	18
154	JI N° 133	18

Fuente: Elaboración propia

Una vez ordenada la distribución debo buscar el Orden de la Mediana. Es la unidad de análisis que divide a la distribución en 2 partes iguales que contiene el valor que divide a la distribución en dos partes iguales. Tradicionalmente se simboliza de este modo $Md(o)$.

Cálculo del Orden de la Mediana:

$$\frac{N}{2} = Md(o)$$

Si el Valor de N es impar utilizar:

$$\frac{N+1}{2} = Md(o)$$

En nuestro ejemplo el Orden de la Mediana es:

$$\frac{154}{2} = 77$$

Si observamos la distribución de frecuencias de la tabla 1.7, vemos que la unidad de análisis N° 77 tiene el valor 4, por lo tanto la Md es **4 años de graduados**

Tabla 1.8: Distribución de frecuencias de la variable Años de Graduado de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Años de graduado	Frecuencia	Frecuencia acumulada	Porcentaje	Porcentaje acumulado
1	27	27	17,5	17,5
2	21	48	13,6	31,2
3	15	63	9,7	40,9
4	21	84	13,6	54,5
5	14	98	9,1	63,6
6	15	113	9,7	73,4
7	5	118	3,2	76,6
8	11	129	7,1	83,8
9	4	133	2,6	86,4
10	2	135	1,3	87,7
11	5	140	3,2	90,9
12	1	141	0,6	91,6
13	5	146	3,2	94,8
14	2	148	1,3	96,1
15	2	150	1,3	97,4
16	2	152	1,3	98,7
18	2	154	1,3	100,0
Total	154		100,0%	

Fuente: Elaboración propia

Cuando una distribución de frecuencias está agrupada, debemos buscar el $Md(o)$ en las frecuencias acumuladas.

Tomando el ejemplo de la tabla 1.8, observamos que la unidad N° 77 se encuentra en categoría 4 años de graduado, por lo tanto este es el valor Md . Bajo el supuesto que dentro de cada categoría las unidades de análisis se distribuyen homogéneamente, podemos utilizar el siguiente cálculo para determinar con mayor precisión el valor de Md .

Cálculo de la Mediana para variables agrupadas:

$$lm + \frac{(Md(o) - faa)}{f} * A = Md$$

lm = Límite inferior del intervalo o categoría de variable

faa = Frecuencia acumulada del intervalo/categoría anterior

f = Frecuencia de la categoría/intervalo del $Md(o)$.

A = Amplitud del intervalo

Siguiendo la distribución de la Tabla 1.8:

$$4 + \frac{(77 - 63)}{84} * 1 = Md$$

$$4 + 0,17 * 1 = Md$$

$$4,17 = Md$$

Tabla 1.9: Distribución de frecuencias de la variable Años de Graduado de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005

Años de graduado	Punto Medio	Frecuencia	Frecuencia acumulada	Porcentaje	Porcentaje acumulado
1 - 2	1,5	48	48	31,2	31,2
3 - 5	4,0	50	98	32,5	63,6
6 - 10	8,0	37	135	24,0	87,7
11- 15	13,0	15	150	9,7	97,4
16 - 20	23,0	4	154	2,6	100,0
Total		154		100,0%	

Fuente: Elaboración propia

Siguiendo la distribución de la Tabla 1.9:

$$3 + \frac{(77 - 48)}{98} * 3 = Md$$

$$3 + 0,35 * 3 = Md$$

$$4,04 = Md$$

VI. Medidas de posición

Las medidas de posición son puntuaciones que separan una fracción de los casos de una distribución. Entre las medidas de posición podemos encontrar a los cuartiles, los deciles y los percentiles.

- *Cuartiles*: dividen a una distribución en cuatro partes iguales. Cada una de estas partes representa el 25% de los casos en cada grupo.
- *Deciles*: dividen a una distribución en diez partes iguales. Cada una de estas partes representa un 10% de los casos en cada grupo.
- *Percentiles*: dividen a una distribución en 100 partes iguales. Según Ritchey (2001), los percentiles son “el porcentaje de casos que caen en o están debajo de un valor específico de X”.

El procedimiento es similar al del cálculo de la mediana. Primero se debe calcular el orden del Cuartil, Decil o Percentil y luego ver el valor o categoría de variable que asume dicha unidad de análisis.

Cálculo de Orden del Cuartil:

$$\frac{N}{4} * n^{\circ} \text{ cuartil} = C(o)$$

Cálculo de Orden del Decil:

$$\frac{N}{10} * n^{\circ} \text{ decil} = D(o)$$

Cálculo de Orden del Percentil:

$$\frac{N}{100} * n^{\circ} \text{ percentil} = P(o)$$

Tomando el ejemplo de la tabla 1.7, para analizar la distribución de frecuencias de acuerdo a sus cuartiles, deciles o percentiles observamos que los valores son:

Tabla 1.10: Valores que asume la variable Años de Graduado de los jóvenes investigadores de la Universidad de Buenos Aires en el año 2005 de acuerdo a cada una de las medidas de posición

Cuartil Nº	Orden del:		Ud. A. Nº:	Valor de la variable Años del graduado	
	Decil Nº	Percentil Nº			
1	1	1	2	1	
		2	4	1	
		3	5	1	
		4	7	1	
		5	8	1	
		6	10	1	
		7	11	1	
		8	13	1	
		9	14	1	
		10	16	1	
	2	2	11	17	1
			12	19	1
			13	21	1
			14	22	1
			15	24	1
			16	25	1
			17	27	1
			18	28	2
			19	30	2
			20	31	2
	3	3	21	33	2
			22	34	2
			23	36	2
			24	37	2
			25	39	2
			26	41	2
			27	42	2
			28	44	2
			29	45	2
			30	47	2
2	4	31	48	2	
		32	50	3	
		33	51	3	
		34	53	3	
		35	54	3	
		36	56	3	
		37	57	3	
		38	59	3	
		39	61	3	
		40	62	3	
	5	5	41	64	4
			42	65	4
			43	67	4
			44	68	4
			45	70	4
			46	71	4
			47	73	4
			48	74	4
			49	76	4
			50	77	4

Orden del:			Ud. A. Nº:	Valor de la variable Años del graduado
Cuartil Nº	Decil Nº	Percentil Nº		
3	6	51	79	4
		52	81	4
		53	82	4
		54	84	4
		55	85	5
		56	87	5
		57	88	5
		58	90	5
		59	91	5
		60	93	5
	7	61	94	5
		62	96	5
		63	98	5
		64	99	6
		65	101	6
		66	102	6
		67	104	6
		68	105	6
		69	107	6
		70	108	6
	8	71	110	6
		72	111	6
		73	113	6
		74	114	7
		75	116	7
		76	118	7
		77	119	8
		78	121	8
		79	122	8
		80	124	8
9	81	125	8	
	82	127	8	
	83	128	8	
	84	130	9	
	85	131	9	
	86	133	9	
	87	134	10	
	88	136	11	
	89	138	11	
	90	139	11	
4	10	91	141	12
		92	142	13
		93	144	13
		94	145	13
		95	147	14
		96	148	14
		97	150	15
		98	151	16
		99	153	18
		100	154	18

Bibliografía

- Babbie, E. (1996), Manual para la práctica de la investigación social, Bilbao, Desclée De Bouwer International.
- Cea D'Ancona, M. A. (1996), Metodología cuantitativa. Estrategias y técnicas de investigación social, Madrid, Síntesis.
- Creswell, J. W. (1994), Research Design. Qualitative and Quantitative Approaches, Thousand Oaks, Sage.
- García Ferrando, M. (1992), Socioestadística. Introducción a la estadística en sociología, Madrid, Alianza.
- Holguín Quiñones, F. (1988), Estadística descriptiva aplicada a las ciencias sociales, México, Universidad Autónoma de México.
- Ritchey, F. (2001), Estadística para las ciencias sociales. El potencial de la imaginación estadística, México D.F., McGraw-Hill/Interamericana Editores.