



Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Sociología

Documento de Cátedra 68: Guía de ejercicios de estadística

**Autores (2010): Otamendi, María Alejandra; Ballesteros, Matías;
Paredes, Diego y Almendros, Teresa.**

(incluye ejercicios de la Guía elaborada por Martínez Mendoza, Ramiro y Griselli, Lucia, 2007).

Revisión 2012

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.

Estos documentos son material de uso interno y no pueden ser incorporados a ediciones impresas ni reproducidos comercialmente.

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

LA MATRIZ DE DATOS: VARIABLES Y CATEGORÍAS

El departamento de Investigación Educativa de la Universidad XX está interesado en estudiar el problema de la enseñanza y el aprendizaje de la lectura en el EGB.¹ Se trata de encontrar el método o los métodos de enseñanza de la lectura que mejor se adapten a alumnos varones, que son las unidades de análisis de este estudio. Son muchos los factores que influyen en las dificultades de aprendizaje lector, por ejemplo: biológicos, sociales, psicológicos, y educativos.

Para el presente estudio las variables implicadas se definen de la siguiente manera:

Nº: Número de identificación de las unidades de análisis

V1: Método de enseñanza de la lectura:

El estudio incluye tres métodos, cada uno de los cuales representa un enfoque teórico- didáctico muy diferente

- 1) método sintético
- 2) método global
- 3) método mixto

V2: Asistencia al preescolar

- 1) sí asistió
- 2) no asistió

V3: Madurez lectora

Puntuación total obtenida por un sujeto en un test ABC, dicha puntuación oscila entre 0 y 20 puntos

V4: Ambiente familiar

Dado por el nivel de interacción del niño en el seno de la familia

- 1) mayor interacción en el seno de la familia
- 2) media interacción en el seno de la familia
- 3) menor interacción en el seno de la familia

V5: Nivel socioeconómico

- 1) alto
- 2) medio
- 3) bajo

V6: Nivel de vocabulario previo al entrar al EGB

La prueba utilizada fue el test de conceptos básicos. El objetivo del test es medir tanto los preconceptos como el vocabulario en uso. Varía entre 0 y 30 puntos

V7: Nivel intelectual

Se midió inteligencia general a través de un test de inteligencia. Varía entre 0 y 43 puntos.

V8: Velocidad lectora

La puntuación varía entre 0 y 70 puntos

V9: Memoria general

Se utilizó una escala que mide la capacidad de memorizar contenidos. La puntuación máxima es de 37 puntos.

V10: Comprensión lectora

Test que mide comprensión, exactitud y velocidad lectora. Puntuación máxima 30 puntos.

V11: Cálculo

Se utilizó una prueba de cálculo que mide básicamente cálculo aritmético y es de aplicación colectiva. La puntuación máxima es de 28 puntos.

V12: Estilo cognitivo

¹ La matriz de datos utilizada para algunos ejercicios fue extraída de: Fernández Díaz; García Ramos; Asensio Muñoz y Fuentes Vicente (1990) Problemas de estadística aplicados a las ciencias sociales, Madrid, Ed. Síntesis.

La puntuación máxima que podía obtenerse es de 18 puntos. Se aplicó un test que en el que se les pedía a los niños que reconocieran figuras simples de entre un conjunto de 18 figuras complejas.

V13: Adaptación escolar

La puntuación máxima es 10 = máxima adaptación y la mínima es 1= mínima adaptación.

V14: Nivel de aspiraciones de los padres

Aspiraciones con respecto a la formación de sus hijos, especialmente acerca del nivel de estudios que pretenden que sus hijos alcancen. Medido a través de la pregunta: ¿Se ha planteado que su hijo termine estudios superiores?

- 1) sí
- 2) no

Modelo de análisis propuesto. Orden temporal de las variables

I	II	III	IV ⁽²⁾
Nivel socioeconómico	Ambiente familiar	Asistencia al preescolar	Madurez lectora
	Nivel intelectual ⁽¹⁾	Método de enseñanza	Nivel de vocabulario
	Aspiraciones de los padres		Velocidad Lectora
			Memoria general
			Cálculo
			Estilo cognitivo
			Adaptación escolar

(1) El nivel intelectual se desarrolla en el hogar y en el ámbito escolar

(2) Existe interacción entre las variables

Matriz de datos

Nº	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
1	1	1	18	1	1	28	43	69	30	29	28	17	9	1
2	1	1	17	1	2	27	30	68	20	30	22	16	8	1
3	1	1	7	2	3	14	18	38	30	10	15	9	6	1
4	1	1	12	2	2	23	23	50	26	21	25	14	5	1
5	1	1	15	3	1	24	19	57	37	24	19	8	3	1
6	1	1	6	3	3	14	22	33	15	11	20	11	4	2
7	1	2	7	1	3	14	26	30	21	12	27	13	7	1
8	1	2	10	1	2	18	20	38	27	8	19	12	6	2
9	1	2	9	2	1	14	19	39	20	7	16	10	4	2
10	1	2	4	2	3	10	18	22	32	5	12	8	5	2
11	1	2	2	3	2	5	12	20	16	3	9	4	1	2
12	1	2	7	3	1	14	17	37	27	6	14	9	3	2
13	2	1	20	1	1	30	29	62	30	28	25	15	7	1
14	2	1	9	1	3	18	22	35	30	12	21	11	8	2
15	2	1	10	2	2	15	20	41	25	9	24	6	7	1
16	2	1	13	2	2	20	25	50	22	20	22	16	4	1
17	2	1	11	3	1	16	15	43	21	16	11	8	2	2
18	2	1	2	3	3	8	11	19	12	2	8	10	1	2
19	2	2	16	1	1	22	33	55	29	28	26	18	5	2
20	2	2	8	1	2	18	28	30	16	11	19	16	7	1
21	2	2	3	2	3	7	24	24	10	4	23	12	5	2
22	2	2	9	2	2	12	20	47	19	7	20	7	2	1
23	2	2	4	3	3	9	13	21	30	5	10	5	4	2
24	2	2	2	3	1	8	14	23	22	3	12	6	1	2
25	3	1	19	1	2	27	27	68	34	25	21	13	7	1
26	3	1	14	1	3	24	22	60	32	23	16	8	6	1
27	3	1	15	2	1	24	28	70	30	22	18	13	4	1
28	3	1	9	2	3	17	19	31	24	10	16	6	7	1
29	3	1	8	3	2	12	18	28	20	7	13	5	2	2
30	3	1	11	3	1	18	20	46	19	15	17	7	2	2
31	3	2	10	1	3	14	20	50	12	6	14	8	5	1
32	3	2	13	1	1	18	24	48	19	18	20	14	4	2
33	3	2	10	2	2	17	21	37	26	18	19	12	3	1
34	3	2	4	2	3	10	18	35	14	6	10	5	2	1
35	3	2	10	3	1	16	10	42	18	10	7	3	1	1
36	3	2	3	3	2	12	14	17	22	2	9	4	2	2

CONSTRUCCIÓN DE DISTRIBUCIONES DE FRECUENCIAS

1.- ¿Cuál es la unidad de análisis del estudio?

2.- Indique cada una de las variables que componen la matriz de datos y su nivel de medición

3.- Construya las distribuciones de frecuencias de las siguientes variables:

- **Ambiente familiar (V4)**
- **Nivel socioeconómico (V5)**
- **Nivel de vocabulario previo al entrar al EGB (V6)**
- **Nivel intelectual (V7)**

4.- Agrupe en 5 intervalos iguales las distribuciones de frecuencias de las siguientes variables, indicando la amplitud de cada intervalo y los puntos medios:

- **Nivel de vocabulario previo al entrar al EGB (V6)**
- **Nivel intelectual (V7)**

5.- Dada la siguiente distribución de frecuencias de la variable **comprensión lectora (V10)** realizada con el software SPSS:

- a) Agrúpela en 6 intervalos.
- b) Indique la amplitud de cada intervalo y los puntos medios.

Cuadro Nº 1. Comprensión lectora de los alumnos varones que cursan el EGB

	Frequency	Percent	Cumulative Percent
Valid 2	2	5,6	5,6
3	2	5,6	11,1
4	1	2,8	13,9
5	2	5,6	19,4
6	3	8,3	27,8
7	3	8,3	36,1
8	1	2,8	38,9
9	1	2,8	41,7
10	3	8,3	50,0
11	2	5,6	55,6
12	2	5,6	61,1
15	1	2,8	63,9
16	1	2,8	66,7
18	2	5,6	72,2
20	1	2,8	75,0
21	1	2,8	77,8
22	1	2,8	80,6
23	1	2,8	83,3
24	1	2,8	86,1
25	1	2,8	88,9
28	2	5,6	94,4
29	1	2,8	97,2
30	1	2,8	100,0
Total	36	100,0	

Fuente: elaboración propia.

6.a. Indique el nivel de medición de las siguientes variables:

- I. Auto percepción del estado de salud: Excelente, muy bueno, bueno, regular, malo o muy malo.
- II. Consumo de Remedios sin prescripción médica durante los últimos 6 meses: Sí, No.
- III. Peso al nacer: expresado en kilogramos.
- IV. Gasto mensual en salud: expresado en pesos.

6.b. Agrupe los sistemas de categorías para las variables III y IV, modificando su nivel de medición. Justifique el nuevo sistema de categorías.

7.a. En la Ciudad de Buenos Aires en el año 2008 se llevó a cabo un estudio sobre la situación de dependencia de personas de la tercera edad: se estudió en particular a aquellas personas mayores cuyo cuidado se encontraba a cargo de un familiar. Se realizaron entrevistas con el objetivo de conocer las características tanto de la persona mayor cuidada como del cuidador.

Indique el sistema de categorías y el nivel de medición para cada una de las siguientes variables del estudio:

Variabes del cuidador	Variabes de la persona mayor cuidada
Distancia (en cuadras) entre la casa y el trabajo	Sexo
Edad	Peso
Nivel de educación alcanzado	Nivel de severidad de la enfermedad

7.b. Proponga dos variables para medir características del cuidador y dos para medir características de la persona cuidada, con sus respectivos sistemas de categorías, indicando el nivel de medición. Justifique su propuesta.

8.- Señale si el sistema de categorías propuesto para cada una de las variables abajo mencionadas cumple con los criterios de ser exhaustivo y excluyente. Si no es así, corrija el sistema de categorías para cumplir con dichos criterios.

A. Tiempo transcurrido desde la última consulta médica

- Entre 1 y 4 meses.
- Entre 6 meses y 9 meses
- Entre 9 meses y 1 año.

B. Cantidad de consultas realizadas a un profesional de la salud en el último año:

- De 1 a 3
- De 3 a 5
- De 6 a 7
- De 9 a 11

C. Religión de la persona atendida en la sala de urgencias de un hospital:

- Católica
- Protestante
- Judía
- Ninguna

9.- ¿Qué nivel de medición corresponde a la variable “máximo nivel de educación alcanzado” cuando es medido como:

- a) Número de años de estudio realizados (0, 1, 2, 3, ...)
- b) Nivel educativo completado (primario, secundario, universitario, posgrado)

10. – En las siguientes distribuciones de frecuencias calcule los porcentajes y realice la lectura **univariada** de cada cuadro:

Cuadro Nº 2. Presencia de enfermedades crónicas diagnosticadas por un médico

Sí	180
No	971
Total	1151

Fuente: elaboración propia. Datos ficticios.

Cuadro Nº 3. Frecuencia en la compra de medicamentos con receta médica durante el último año	
No compró medicamentos bajo receta	50
Entre 1 y 4 veces	211
Entre 5 y 8 veces	575
Entre 9 y 12 veces	240
Más de 12 veces	75
Total	1151

Fuente: elaboración propia. Datos ficticios.

Cuadro Nº 4. Tipo de establecimiento en el que le recetaron su último medicamento este año	
No le recetaron medicamentos este año	100
Salita o centro de salud	210
Hospital público	430
Clínica privada	311
Consultorio particular	80
Otros	18
NS/NC	2
Total	1151

Fuente: elaboración propia. Datos ficticios.

Cuadro Nº 5. Sexo del paciente	
Varón	572
Mujer	579
Total	1151

Fuente: elaboración propia. Datos ficticios.

11.- Se realizó una encuesta a 639 usuarios de hospitales públicos de la Ciudad de Buenos Aires para conocer su satisfacción con la atención médica recibida. A partir de varias preguntas, se construyó la variable “nivel de satisfacción con la atención médica recibida en el hospital”, con un valor mínimo de 0 y un valor máximo de 100. Los resultados son los siguientes:

Cuadro Nº 6. Frecuencia del nivel de satisfacción con la atención médica recibida por usuarios de hospitales de la Ciudad de Buenos Aires

Nivel de satisfacción con la atención médica recibida	Frecuencias
00-10	11
11-20	25
21-30	35
31-40	47
41-50	70
51-60	116
61-70	141
71-80	113
81-90	59
91-100	22
Total	639

Fuente: Elaboración propia. Datos Ficticios.

Incluya en la tabla los puntos medios de los intervalos, las frecuencias absolutas acumuladas, las frecuencias porcentuales, las frecuencias porcentuales acumuladas.

12.- A continuación se presenta la distribución del parentesco de los familiares cuidadores de personas que sufren Alzheimer. Calcule la razón de mujeres a hombres cuidadores.

Cuadro N° 7. Tipo de parentesco del familiar cuidador de paciente con Alzheimer

Parentesco del familiar con el paciente	Frecuencias
Esposa	114
Esposo	17
Hija	37
Hijo	4
Hermana	8
Hermano	1
Madre	2
Nuera	13
Otros parientes femeninos	9
Otros parientes masculinos	2
TOTAL	207

Fuente: elaboración propia. Datos ficticios.

13.- La Organización Mundial de la Salud (OMS) ha realizado una encuesta, durante el año 2006, dirigida a conocer la opinión de una muestra aleatoria de 2900 médicos de cuatro países distintos. Una de las preguntas del cuestionario se encontraba formulada de la siguiente manera: "En su opinión, ¿cuál es el nivel de familiaridad que tienen los legisladores en salud pública de su país con la Estrategia Global de Salud de la OMS? Los resultados se presentan en el siguiente cuadro.

TITULO

	EEUU	España	Argentina	Rusia
Muy familiarizados	188	149	334	194
Bastante familiarizados	326	241	475	455
Para nada familiarizados	116	151	92	56
No sabe / No responde	53	32	23	17

Fuente: elaboración propia. Datos ficticios.

- A partir de los sistemas de categoría presentados, indique el nombre de las variables del cuadro, señalando su nivel de medición.
- Coloque el título correspondiente.
- Responda a las siguientes preguntas, construyendo las distribuciones de frecuencia necesarias para poder responderlas:
 - ¿Qué porcentaje de encuestados considera que los legisladores están **muy familiarizados** con la Estrategia Global de la OMS?
 - ¿Qué porcentaje de encuestados considera que los legisladores están **para nada familiarizados** con dicha estrategia?
 - Para cada país, ¿Qué porcentaje considera que los legisladores están bastante familiarizados con la estrategia?
 - ¿Cuál es el país con mayor proporción de encuestados que **no sabe o no responde** la pregunta?

14.- Teniendo en cuenta las distintas definiciones de envejecimiento demográfico y los datos que figuran a continuación, calcule e interprete para los distintos años:

- la proporción de personas mayores de 65 años.
- la tasa de mayores de 65 años cada 100.000 habitantes.
- la razón entre la población adulta mayor y joven.

Envejecimiento demográfico (envejecimiento poblacional): proceso por el cual una población llega a ser más vieja, de acuerdo con ciertos criterios, tales como incremento en la proporción de personas mayores de 60 / 65 años, en relación con la población total. Otros criterios son: expectativa promedio de vida, edad promedio de una población, proporción comparada de

población vieja y joven. (En Glosario de la Red Latinoamericana de Gerontología: <http://www.gerontologia.org/portal/glosario.php>).

Cuadro N° 9. Indicadores demográficos seleccionados para fechas censales
Total del país. Años 1869-2010

Años	Porcentaje de		Esperanza de vida al nacer, en años (3)	Población Total (en millones)
	Adultos mayores	Jóvenes (2)		
1869	2,2	42,8	32,9	1,8
1895	2,5	40,3	40,0	4,0
1914	2,3	38,4	48,5	7,9
1947	3,9	30,9	61,1	15,8
1960	5,6	30,8	66,4	20,0
1970	7,0	29,3	65,6	23,3
1980	8,2	30,3	68,9	27,8
1991	8,9	30,6	71,9	32,6
2001	9,9	28,3	73,8	36,2
2010	10,2	25,5	75,5	40,1

(1) Adultos mayores: población de 65 años y más de edad.

(2) Jóvenes: población de 0 a 14 años de edad.

(3) Los valores de 1869 y 1895 corresponden a los dos primeros períodos intercensales, mientras que los valores sucesivos provienen de Tablas de Vida elaboradas para cada fecha censal.

Fuente: INDEC, Censos Nacionales de Población, Hogares y Viviendas.

INDEC (1975), *La población de Argentina*, Buenos Aires, Serie Investigaciones Demográficas N° 1.

INDEC (1988), *Tabla de mortalidad 1980-1981, total y jurisdicciones*. Buenos Aires, Estudios 10.

INDEC (1995), *Tabla completa de Mortalidad de la Argentina por sexo. 1990-1992*. Buenos Aires, Serie Análisis Demográfico N° 3.

INDEC (2004), *Estimaciones y proyecciones de población. Total del país. 1950-2015*, Buenos Aires, Serie Análisis Demográfico N° 30.

Pantelides E.A. (1989), *La fecundidad argentina desde mediados del siglo XX*, Buenos Aires, CENEP, Cuadernos del CENEP N° 41.

CONSTRUCCIÓN DE CUADROS BIVARIADOS

15.- A partir de los datos de la matriz incluida al principio de esta guía, construya los cuadros bivariados correspondientes para comprobar o refutar las hipótesis de trabajo planteadas a continuación, y realice su lectura.

- El nivel socioeconómico familiar incide en el nivel de aspiraciones de los padres de tal manera que a mayor nivel socioeconómico de la familia los padres sí quieren que sus hijos terminen los estudios superiores.
- El ambiente familiar incide sobre el estilo cognitivo, de tal forma que para aquellos niños en cuyo ambiente familiar se da una mayor interacción, el estilo cognitivo es mayor.
- La asistencia al preescolar incide en el estilo cognitivo de los niños, de tal forma que en aquellos niños que han asistido al preescolar el nivel cognitivo es mayor.
- Compare las conclusiones de los cuadros b y c.

16.- Dado el siguiente cuadro bivariado, elaborado utilizando la información de la matriz de datos de alumnos varones del EGB:

- Identifique las variables que intervienen, sus categorías y su nivel de medición.
- Postule una relación o hipótesis de trabajo entre las variables, indique cuál es la variable independiente y cuál la variable dependiente.
- Realice la lectura correspondiente, enuncie si se comprueba o si se refuta la relación planteada. Justifique su respuesta.
- Grafique el cuadro.

Cuadro Nº 10. Nivel de vocabulario previo al entrar al EGB, según asistencia al preescolar

		Asistencia al preescolar		Total
		sí	no	
nivel de vocabulario previo	bajo	5	12	17
		27,8%	66,7%	47,2%
	medio	5	6	11
		27,8%	33,3%	30,6%
	alto	8	0	8
		44,4%	0,0%	22,2%
Total (n)	18	18	36	
	100,0%	100,0%	100,0%	

Fuente: elaboración propia.

17.- El Ministerio de Salud de la Ciudad de Buenos Aires llevó a cabo una investigación sobre la opinión de médicos que trabajan tanto en hospitales públicos como en clínicas privadas de la ciudad durante el año 2009 con respecto a diversos temas de salud. Una de las preguntas fue “¿Cuán de acuerdo está usted con que el Estado subsidie las instituciones de salud privadas?” a partir de la cual se obtuvo el siguiente cuadro:

	Clínicas privadas	Hospitales públicos	Total
Nada de acuerdo	30	51	81
Algo de acuerdo	85	90	175
Muy de acuerdo	110	74	184
Total	225	215	440

- Indique las variables que componen el cuadro (si no están explícitas colóqueles el nombre dentro del cuadro); el nivel de medición de cada una de las variables y sus correspondientes categorías.
- Indique la unidad de análisis.
- Agregue el título y la fuente al cuadro.
- Postule una hipótesis a poner a prueba, indicando el status que cumple en este cuadro cada una de las variables (VD, VI).
- Conclusiones de la lectura del cuadro ¿la hipótesis se refuta o no?

18.- En base al Censo Nacional de Población, Hogares y Viviendas 2010, realizado por el Instituto Nacional de Estadística y Censos (INDEC) se obtuvo el siguiente cuadro sobre la población de la provincia del Chaco.

TITULO

Edad	Chaco		
	Posee plan médico	No posee plan médico	Total
0-14	102.295	217.255	319.550
15-64	274.082	382.030	656.112
65 y más	64.542	7.832	72.374
Total	440.919	607.117	1.048.036

- Identifique las variables del cuadro y sus categorías.
- ¿Qué papel juega cada una de ellas (VD, VI)?
- Coloque el título y la fuente correspondiente al cuadro.
- Calcule los porcentajes en el sentido de la variable independiente.
- Realice una lectura del cuadro y extraiga conclusiones.

19.- A partir del siguiente cuadro sobre la población con necesidades básicas insatisfechas (NBI) de la Provincia de Buenos Aires, responda las siguientes preguntas:

Cuadro 11. Porcentaje de la población con NBI según grupos de edad de la Provincia de Buenos Aires, 2001

Grupos de edad	% con NBI	Total en absolutos
0-14 Años	23,9 %	(3669075)
15-64 Años	13,0 %	(8617116)
65 y más Años	11,2 %	(1422711)
Total	15,8 %	(13708902)

Nota: Base población de la provincia de Buenos Aires

Fuente: Censo Nacional de Población, Hogares y Viviendas 2001, INDEC, Procesado con Redatam+SP

- ¿Cuáles son las variables y sus respectivas categorías en el cuadro?
- ¿Qué datos se muestran y cuáles se ocultan en el cuadro? ¿Por qué ocurre esto?
- ¿En qué sentido están calculados los %? Interprete.
- Complete el cuadro colocando la categoría ausente.
- ¿Qué conclusiones puede extraer comparando la distribución con NBI en cada grupo etario?
- ¿Cuál es el número total de personas con NBI en la provincia de Buenos Aires en el 2001?

20.- A partir de una encuesta acerca de la salud mental en la tercera edad, realizada en la Provincia de Buenos Aires en el año 2008, se intenta poner a prueba la hipótesis que señala que en adultos mayores el nivel de autoestima se encuentra asociada a su condición de actividad,

tendiendo los que son activos a tener un mayor nivel de autoestima que los que no lo son y viceversa.

- Identifique la variable dependiente y la variable independiente, con sus categorías y nivel de medición.
- Señale la unidad de análisis.
- Agregue un título al cuadro.
- A partir del cuadro, compruebe la hipótesis enunciada, extraiga conclusiones.

TITULO

Condición de Actividad (%)	Nivel de autoestima			Total
	Bajo	Medio	Alto	
Económicamente No Activo	27%	32%	41%	(650)
Económicamente Activo	40%	25%	35%	(688)

Fuente: elaboración propia

21.- Dada las variables que configuran el cuadro 12 (estado civil y condición de actividad):

- Formule una hipótesis indicando el rol que juega cada una.
- Calcule los porcentajes en función de dicha hipótesis.
- Realice la lectura del cuadro en función de la hipótesis citando los porcentajes.
- Extraiga las conclusiones pertinentes.

Cuadro 12. Condición de actividad de las mujeres según estado civil (en absolutos)

Estado Civil	Condición de actividad		Total
	Económicamente Activa	Económicamente No Activa	
Soltera	55	73	128
Casada	69	276	345
Unida de hecho	54	180	234
Viuda	10	75	85
Separada / Divorciada	27	18	45
Total	215	622	837

Fuente: Muestra del Censo Nacional de Población y Vivienda de 1991

GRÁFICOS

22.-

- Grafique cada uno de los cuatro cuadros que figuran en el ejercicio 10 de la presente guía (cuadros 2, 3, 4 y 5). Justifique su elección de tipo de gráfico (según el nivel de medición de las variables).
- Repita el ejercicio para el cuadro 6 del ejercicio 11. Justifique el tipo de gráfico elegido.
- Repita el ejercicio para el cuadro 7 del ejercicio 12. Justifique el tipo de gráfico elegido.

NOTA: recuerden todos los elementos que deben estar presentes en los gráficos (título, fuente, etc.)

MEDIDAS DE TENDENCIA CENTRAL Y POSICIÓN

23.- Defina la media, mediana y modo. Especifique las limitaciones y ventajas potenciales de cada medida.

24.- ¿Cómo afecta un sesgo a la izquierda a la media, mediana y modo de una distribución de frecuencias? ¿Y un sesgo a la derecha?

25.- Utilizando la matriz de alumnos varones de EGB, se necesita conocer los valores de la mediana, la moda y la media aritmética de las siguientes variables:

- **V9: Memoria general**
- **V12: Estilo cognitivo**
- **V6: Nivel de vocabulario previo al entrar al EGB.** Para esta variable realice los cálculos utilizando la matriz de datos y el cuadro confeccionado en el ejercicio 3.

26.- Deseamos saber la media aritmética de la muestra de 36 alumnos en la **variable madurez lectora** a partir de las medias que han obtenido en las tres submuestras de la **variable métodos de enseñanza**:

- **el que siguió el método 1: media 9,50 puntos**
- **el que siguió el método 2: media 8,92 puntos**
- **el que siguió el método 3: media 10,50 puntos**

27.- Supongamos que la prueba de madurez lectora la hubiéramos aplicado además a otras tres muestras de 40, 45 y 55 sujetos y hubiéramos obtenido respectivamente las siguientes medias: 14,8; 10 y 12,8 ¿Cuál sería en este caso la media total?

28.- ¿Cuál es el valor del Cuartil 3, Decil 6 y Percentil 45 de la variable madurez lectora?

29.- Se relevaron las edades de los pacientes adultos mayores que recibió el Hospital Ramos Mejía durante la primera semana del mes de noviembre del año 2008.

- a) Indique la unidad de análisis.
- b) Realice un **análisis univariado** utilizando las medidas de tendencia central.
- c) Calcule el valor del decil 8, percentil 30 y 70. Formule conclusiones.
- d) Reagrupe la variable edad siguiendo el criterio teórico expuesto a continuación.
- e) Calcule la moda, la media y la mediana una vez agrupados los datos. ¿Qué diferencias encuentra con los resultados del punto b? ¿Cuáles son las ventajas y los inconvenientes de agrupar los datos?

Clasificación teórica:

“Hoy se pueden delimitar tres subgrupos, en orden a criterios de actividad económica, intelectual-cognoscitivo, cultural, de salud y de esperanza de vida a esas edades. Los umbrales son flexibles, aunque dependen de esas variables y hay problemas claros de denominación. En una primera aproximación cabe distinguir:

1. **Vejez inicial o incipiente**, entre los 60 y 69 años de edad, en que aún hay muchas personas activas que no han salido del mercado de trabajo y conservan en general buenas condiciones de salud.
2. **Vejez intermedia**, entre el 70 y el 84 aniversario, en que ya casi todas las personas han salido del mercado de trabajo, salvo ciertos trabajadores por cuenta propia: agricultores, empresarios, intelectuales, etc., que guardan ciertas condiciones de salud mental y física y que observan y perciben la muerte aún como un fenómeno no obligatoriamente próximo.
3. **Vejez avanzada**, de más de 85 años de edad, que corresponde a las personas al final de su línea o ciclo de vida”.

Fuente: Martín Ruiz, J. F. (2005) Los factores definitorios de los grandes grupos de edad de la población: tipos, subgrupos y umbrales. *Revista electrónica de geografía y ciencias sociales*, Vol. IX, núm. 190.

Cuadro 13. Distribución de frecuencia de la variable edad de los adultos mayores atendidos en el Hospital Ramos Mejía durante la primera semana de Noviembre, 2008.

Edad	Frecuencia	Porcentaje*	Porcentaje acumulado
60-64 Años	702	26,7	26,7
65-69 Años	471	17,9	44,6
70-74 Años	502	19,1	63,7
75-79 Años	468	17,8	81,5
80-84 Años	268	10,2	91,7
85-89 Años	125	4,8	96,5
90-94 Años	65	2,5	99,0
95 y más Años	25	0,9	99,9
Total	2626	100	

*Nota: el cálculo no es exacto debido a redondeo.
Fuente: elaboración propia

30.- Suponga que se realizó una encuesta en Octubre del 2010 en La Matanza para determinar el gasto en medicamentos de los residentes de la misma. Se publican los resultados indicando que los menores de 65 años tenían un gasto promedio de \$125 por mes, mientras que quienes tenían 65 años o más gastaban un promedio de \$210 por mes.

Usted necesita saber cuánto es el gasto promedio de todos los residentes de La Matanza, pero esos datos no están disponibles. Sin embargo, cuenta con los siguientes datos del Censo 2010:

Cuadro 14 . Habitantes por edad Agrupada. La Matanza, 2010.

Edad Agrupada	Frecuencia
0-14 Años	475.036
15-64 Años	1.152.051
65 y más Años	148.729
Total	1.775.816

Fuente: Censo Nacional de Población, Hogares y Viviendas 2010. Instituto Nacional de Estadística y Censos (INDEC).

Según estos datos, ¿cuánto fue el gasto promedio en medicamentos de los habitantes de La Matanza en el mes de Octubre del 2010?

31.- La media de las edades de un grupo de 20 personas es de 22 años, y la de otro grupo de 15 personas es de 26 años. ¿Cuál es la media de ambos grupos?

32.- Se realizó una muestra aleatoria de 20 hogares de la Ciudad de Tandil en Mayo del 2010. Se le consultó al Jefe de hogar la cantidad de veces que en el último año visitó a un médico. Las respuestas fueron las siguientes:

2, 1, 2, 4, 2, 3, 4, 2, 1, 4, 2, 1, 0, 3, 2, 4, 3, 4, 2, 2

- Compile los datos en una tabla de distribución de frecuencias
- Indique la unidad de análisis, el nombre de la variable y su nivel de medición
- Realice el análisis univariado, utilizando las medidas de tendencia central.
- Realice el análisis tomando los cuartiles de la distribución.

MEDIDAS DE VARIABILIDAD

33.- La media para una muestra de conjunto de datos equivale a 80. ¿Cuál de los siguientes valores es un valor imposible para la desviación estándar? ¿Por qué? 200, 0, -20

34.- Utilizando la matriz de alumnos varones del EGB indique:

a. ¿Cuál es el rango o amplitud total de la variable madurez lectora en la muestra n=36?

b. ¿Cuál es la desviación estándar de la variable madurez lectora?

35.- Los siguientes son los coeficientes de variación de la muestra de los 36 alumnos varones en las siguientes variables:

Comprensión lectora (V10): 65,60

Velocidad lectora (V 8): 37,31

Según estos valores: ¿Cuál es la variable más homogénea? ¿Por qué?

36.- Dados los siguientes valores obtenidos de las medias y los desvíos estándar, calculados a partir de la matriz de alumnos varones de EGB, para las variables **Nivel intelectual (V 7)** y **Cálculo (V11)**, Indique cuál de estas distribuciones es la más heterogénea. Justifique su respuesta.

Cuadro 15. Medias y desvíos estándar para las diversas variables

Variable	N	\bar{X}	DS
Nivel intelectual	36	21.17	6,64
Cálculo	36	17,42	5,78

37.- A partir de las estadísticas vitales recopiladas por el GCBA, obtenemos la edad promedio de las muertes por SIDA de varones y mujeres en el año 2008. Indique cuál subgrupo es más homogéneo en la edad de muerte por SIDA.

Cuadro 16. Edad promedio de las muertes por SIDA, por sexo. Ciudad de Buenos Aires, 2008

	Media	Desvío estándar*
Varones	44,1	3,2
Mujeres	40,8	3,0

Fuente: Dirección General de Estadística y Censos, Ministerio de Hacienda, GCBA. Estadísticas vitales. * datos ficticios.

38.- Con el fin de estimar los gastos anuales, se hizo un relevamiento para conocer la cantidad de días que los pacientes pasaban internados en la unidad de terapia intensiva (UTI) antes de transferirlos a otra sala del hospital Zubizarreta de la Ciudad de Buenos Aires.

Se le pide que realice un análisis univariado, calculando la moda, la mediana, la media aritmética, y el desvío standard.

Cuadro 17. Distribución de frecuencias del tiempo de internación (en días a la fecha del relevamiento) de los pacientes de la UTI del hospital Zubizarreta de la Ciudad de Bs. As., 2004

Días de internación	Frecuencia
1	34
2	23
3	45
4	50
5	31
6	34
7	22
8	9
9	6
10	2
N	256

Fuente: Elaboración propia. Datos ficticios.

39.- En una muestra de pacientes del Servicio de traumatología del Hospital de Clínicas, la media aritmética de la edad para el total de 20 mujeres es de 50 años y para el total de 20 varones es igual a 30 años. La desviación estándar de los 40 pacientes considerados conjuntamente es igual a 10. Calcular el coeficiente de variación para el total de los pacientes.

40.- En un conjunto de 150 alumnos, la media obtenida en la prueba de evaluación fue de 60 puntos y el desvío de 5 puntos. Si a cada uno de los puntajes originales se les suma 10 puntos:

- ¿Cuál es el valor de la media de la nueva distribución de puntajes?
- ¿Cuál es el desvío estándar de la misma distribución?
- ¿Qué sucede con el coeficiente de variación al sumarle los puntajes?

41.- Una empresa lleva a cabo un estudio de la cantidad de kilómetros recorridos usando transporte público por sus empleados en un día típico. Una muestra aleatoria de diez empleados arroja los siguientes valores (en km.)

0,0,4,0,0,0,10,0,6,0

- Calcule e interprete la media, mediana, modo, rango, varianza y desviación estándar para esos valores.
- Una nueva persona incluida en la muestra vive en una ciudad distinta y viaja 90 kilómetros por día en transporte público. Vuelva a calcular la media, mediana, y desviación estándar, y describa el efecto de la observación extrema.

42.- A mediados de 1980 la Encuesta Social General preguntó a sus encuestados cuántos amigos cercanos tenían. Para una muestra de tamaño 1467, la media fue 7,4 y la desviación estándar fue 11,0. La distribución tenía una mediana de 5 y un modo de 4. Basándose en estos estadísticos, ¿qué puede ud. conjeturar acerca de la forma de la distribución? ¿Por qué?

43.- Para cada caso, decida cuál es el conjunto de puntuaciones más homogéneo:

- C1: $\bar{X} = 5$, $S = 1$
C2: $\bar{X} = 5$, $S = 2$
- C1: $\bar{X} = 5$, $S = 1$
C2: $\bar{X} = 0,01$, $S = 1$
- C1: $\bar{X} = 128$, $S = 5$
C2: $\bar{X} = 4$, $S = 3$
- C1: $\bar{X} = 1.200$, $S = 58$
C2: $\bar{X} = 1.000$, $S = 58$

CURVA NORMAL

44.- Utilizando los datos de la matriz de alumnos varones del EGB, suponga que la distribución de las puntuaciones en la variable Velocidad lectora (V8) siguiera la distribución normal. Con una Media de V8: 41,19 puntos y un Desvío estándar de V8: 15,37

- ¿Qué puntuación habría obtenido un sujeto que tiene debajo de él al 30 % de casos?
- ¿Qué porcentaje de sujetos ha obtenido una puntuación superior a 26 puntos?
- ¿Entre qué puntajes se encuentra el 90% central de la distribución?
- ¿Cuál es la probabilidad de obtener un puntaje superior a 67 puntos?

45.- Un problema central en la salud de los mayores es el nivel de colesterol en sangre porque puede producir un infarto o derrame cerebral. Se considera que el valor deseable es de 200 mg/dl. Una encuesta realizada entre enfermos coronarios indicó que el colesterol promedio era de 240 mg/dl con una desviación standard de 30 mg/dl. Suponiendo una distribución normal para la variable nivel de colesterol calcule:

- Porcentaje de enfermos coronarios por encima del valor de colesterol deseable
- Porcentaje de los enfermos con niveles de colesterol entre 180 y 200 mg/dl.
- Nivel tope correspondiente al 30% de los enfermos coronarios con menor nivel de colesterol.

46.- Según el Censo de 2009 en las Villas 31 y 31bis² el promedio de edad de los habitantes es de 23,3 años (casi 16 años más jóvenes que en el resto de la Ciudad) con una desviación standard de 3 años. Suponga que la distribución de las edades es normal:

- ¿Cuál es el percentil de un residente de 35 años?
- ¿Cuál es la edad del residente que ocupa el percentil 68?

47.- En una encuesta del CEDOP-UBA del 2005, se obtuvo una muestra de 3000 familias y se midió la cantidad de dinero gastado anualmente en medicamentos (datos ficticios). La variable se distribuyó en forma normal. La media es de \$235 y la desviación standard de \$80.

- La familia Rodríguez destina \$220 a la compra de medicamentos ¿Cuántas familias de la muestra destinan a este rubro más dinero que la familia Rodríguez?
- ¿Cuál es el gasto tope en medicamentos por el 70% inferior de la muestra?

48.- La Encuesta Social General de 1990 preguntó a los encuestados, “¿Durante los últimos 12 meses, cuántas personas que usted conocía personalmente fueron víctimas de homicidio?” El siguiente cuadro corresponde al análisis de las respuestas de 1370 personas.

Cuadro 18. Frecuencia de personas que conocen personalmente víctimas de homicidios, 1990

Víctimas	Frecuencia	Porcentaje
0	1244	90,8
1	81	5,9
2	27	2,0
3	11	0,8
4	4	0,3
5	2	0,1
6	1	0,1

Fuente: Encuesta Social General de 1990, EEUU, datos seleccionados.

N	Media	Desv. Est.	100% máx	75% Q3	50% Med	25% Q1	0% Mín
1370	0,146	0,546	6	0	0	0	0

- Calcule la distribución de frecuencias relativas.
- Construya el histograma correspondiente. ¿La distribución es acampanada, sesgada a la derecha, o sesgada a la izquierda?
- Calcule la media, mediana y modo e interprete sus valores.
- Calcule y analice el desvío estándar ¿se aplica la regla empírica a esta distribución, por qué?

² Ver http://estatico.buenosaires.gov.ar/areas/hacienda/sis_estadistico/villa_31_y_31_bis.pdf