



Universidad de Buenos Aires
Facultad de Ciencias Sociales
Carrera de Sociología

Documento de Cátedra 81: Guía de Repaso de Estadística de Metodología de la Investigación Social I

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

GUÍA DE REPASO DE ESTADÍSTICA DE METODOLOGÍA I

1. Como socióloga/o se te encarga dirigir un estudio en esta Facultad para analizar los factores relacionados con un alto rendimiento académico. A partir del listado de las variables que analizarás en el estudio y que figuran en el siguiente cuadro:

- a) señalará qué categorías utilizarías para su medición;
- b) indicará tipo de variable (cualitativas o cuantitativas) y nivel de medición (nominales, ordinales, intervalares o de razón);
- c) indicará las medidas de tendencia central que se pueden utilizar para describir la distribución de cada variable en la población teniendo en cuenta su nivel de medición.

<u>VARIABLES</u>	<u>CATEGORÍAS</u>	<u>TIPO DE VAR. Y NIVEL DE MEDICIÓN</u>	<u>MEDIDAS DE T. CENTRAL</u>
Sexo			
Edad (años cumplidos al momento de la encuesta)			
Año de ingreso a la Universidad (consignar año)			
Cantidad de materias que aprueba por año (en promedio)			
Cantidad de materias aprobadas al momento de la encuesta			
Promedio de notas obtenidas			
Cantidad de horas diarias destinadas a estudiar			
Estado civil			
Condición de actividad			

2. Turner (1995) investigó los efectos del desempleo. Contactó a 5.612 personas, a quienes consideró elegibles para el estudio porque habían estado desempleadas por lo menos una vez, desde que se incorporaron a la fuerza laboral. De estas personas elegibles, realmente entrevistó a 3.617, entre las cuales, 1.252 se integraron en un estudio a largo plazo (de los cuales 732 eran hombres y 520 eran mujeres). En el grupo de estudio a largo plazo, 154 estuvieron "recientemente desempleadas", pues perdieron sus trabajos en los últimos tres años; de ellas, 45 permanecían desempleadas.

- a) ¿Qué proporción de los sujetos elegibles fue entrevistada?
 - b) ¿Qué porcentaje de los que fueron entrevistados no participó en el estudio a largo plazo?
 - c) De los que sí participaron ¿qué porcentaje quedó desempleado en los últimos tres años?
 - d) ¿Qué porcentaje de los recientemente desempleados volvió a trabajar?
 - e) ¿Qué razón de varones y mujeres hay entre los participantes que se integraron al estudio de largo plazo?
3. Una empresa de selección de personal necesita cubrir varios cargos de vendedores. A la prueba de fluidez verbal se presentan 547 candidatos obteniéndose los siguientes resultados.

Tabla N° 1: Distribución de los candidatos según calificación obtenida en la prueba de fluidez verbal

Calificación	Frecuencia simple
0 - 10	8
11 - 20	22
21- 30	32
31- 40	59
41- 50	67
51- 60	120
61-70	131
71- 80	60
81- 90	39
91- 100	9
Total	547

Fuente: datos ficticios

- a) Incorporará a esta tabla el punto medio de cada intervalo, las frecuencias absolutas acumuladas, las frecuencias porcentuales y las frecuencias porcentuales acumuladas.
 - b) Representará gráficamente por medio de un histograma la distribución de esta variable y agregará el título al gráfico. ¿Qué forma adquiere esta distribución?
 - c) Calculará las medidas de tendencia central (moda, mediana y media), de posición (todos los cuartiles, decil 8, percentil 64) y de dispersión (varianza, desvío estándar) para resumir la información contenida en la tabla. Interpretará los resultados.
 - d) Agrupará a los candidatos en 3 niveles de calificación (bajo, medio y alto), asegurándose de que los intervalos tengan igual rango/amplitud. Realizará la lectura univariada de nivel de calificación.
4. En la Encuesta Social General de 1990 (EEUU.) se les preguntó a los encuestados: "Durante los últimos 12 meses, ¿cuántas personas que Ud. conocía personalmente

fueron víctimas de homicidio?” La siguiente tabla muestra un primer reporte del análisis de las respuestas de 1.370 sujetos.

Tabla N° 2: Distribución de la población según cantidad de víctimas conocidas

Víctimas	Frecuencia simple	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
Total	1.370	100,0

Fuente: Encuesta Social General (EEUU.), 1990.

N	Media	Desvío E.	100% Max	75% Q3	50% Mediana	25% Q1	0% Min
1370	0.146	0.546	6	0	0	0	0

- Informá la distribución de las frecuencias relativas.
- Dibujá un histograma. ¿Es una distribución acampanada, asimétrica hacia la derecha o asimétrica hacia la izquierda?
- Interpretá la media, mediana y moda.
- Informá e interpretá el desvío estándar.

5. Sobre la base del siguiente conjunto de datos muestrales:

Tabla N° 3: Cantidad de veces que varones y mujeres donaron sangre en el año

Varones		Mujeres	
Número de donaciones anuales	Frecuencia simple	Número de donaciones anuales	Frecuencia simple
0	23	0	36
1	36	1	27
2	18	2	21
3	14	3	18
4	9	4	2
5	6	5	6
6	2	6	7
Total	108	Total	117

Fuente: datos ficticios

- Calculá la medida de tendencia central apropiada para determinar quiénes donan sangre con mayor frecuencia, si los hombres o las mujeres.
- A partir de estos datos, determiná cuál de los grupos tiene una variabilidad mayor en la cantidad de sangre donada. Tenga en cuenta que los varones tienen un desvío de 1.58 y las mujeres de 1.75. Comentá los resultados que obtuviste.
- Calculá la media conjunta de donación de sangre de varones y mujeres.
- Calculá el rango y el rango intercuartílico.

6. En la ciudad de Bahía Blanca se está por realizar una elección interna abierta para elegir el próximo candidato a intendente del PJ y se desea saber cuál va a ser el grado de participación de los ciudadanos en las mismas. Para ello se realizó una encuesta por muestreo en la que se obtuvieron los siguientes resultados.

Tabla N° 4: Intención de votar en la próxima interna del PJ según condición socioeconómica de los ciudadanos

Intención de votar en la próxima interna	Condición socioeconómica		Total
	Pobres	No pobres	
SI	150	800	950
NO	250	500	750
Total	400	1300	1700

Fuente: datos ficticios

En los resultados muestrales:

- ¿Qué porcentaje de no pobres tienen la intención de participar en la interna?
- ¿Qué porcentaje de pobres tienen la intención de participar en la interna?
- ¿Cuántos no pobres no tienen la intención de participar?
- ¿Cuántos de los encuestados dijeron que sí irán a votar?
- ¿Existe una relación entre condición socioeconómica e intención de voto?
- En esta muestra, ¿cuál es el porcentaje de pobres?

7. A partir del siguiente cuadro:

Tabla N° 5: (completar el título)

Nivel de expectativas de emigración	Años de escolaridad				Total
	0	1 a 7	8 a 12	Más de 12	
Alto	30	40	32	60	162
Medio	40	52	20	30	142
Bajo	58	75	18	20	171
Total	128	167	70	110	475

Fuente: datos ficticios

- a) Inventá un título para el cuadro que contenga todos los elementos necesarios.
 - b) Indicá las variables que contiene el cuadro y la unidad de análisis. Delimitá conceptualmente un universo de estudio del cual se podría haber extraído esta muestra.
 - c) Planteá una hipótesis que relacione las variables. Calculá los porcentajes de modo tal que pueda contrastarse tu hipótesis.
 - d) Realizá la lectura del cuadro y extraé conclusiones.
 - e) Identificá el nivel de medición de las variables del cuadro. Señalá si es posible calcular la mediana y el promedio de años de escolaridad tal como están presentados los datos en el cuadro. Justificá tu respuesta.
8. En una encuesta del CEDOP-UBA del año 2005, se obtuvo una muestra de 3000 familias y se midió la cantidad de dinero gastado anualmente en medicamentos.¹ La variable se distribuyó en forma normal. La media es de \$235 y la desviación standard de \$80.
- a) 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?
 - b) ¿Cuál es la cantidad máxima de dinero gastado en este rubro por el 70% inferior de la muestra?
 - c) ¿Cuántas familias gastan entre \$300 y \$400 pesos en medicamentos?
 - d) ¿Más de cuánto dinero gasta en medicamentos las 30 familias que más gastan?

NOTA: recuerden dibujar la curva normal. A continuación tienen la tabla con las áreas bajo la curva normal.

¹ Datos ficticios.

Áreas bajo la curva normal

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,000000	0,003989	0,007978	0,011967	0,015953	0,019939	0,023922	0,027903	0,031881	0,035856
0,1	0,039828	0,043795	0,047758	0,051717	0,055670	0,059618	0,063559	0,067495	0,071424	0,075345
0,2	0,079260	0,083166	0,087064	0,090954	0,094835	0,098706	0,102568	0,106420	0,110261	0,114092
0,3	0,117911	0,121719	0,125516	0,129300	0,133072	0,136831	0,140576	0,144309	0,148027	0,151732
0,4	0,155422	0,159097	0,162757	0,166402	0,170031	0,173645	0,177242	0,180822	0,184386	0,187933
0,5	0,191462	0,194974	0,198468	0,201944	0,205402	0,208840	0,212260	0,215661	0,219043	0,222405
0,6	0,225747	0,229069	0,232371	0,235653	0,238914	0,242154	0,245373	0,248571	0,251748	0,254903
0,7	0,258036	0,261148	0,264238	0,267305	0,270350	0,273373	0,276373	0,279350	0,282305	0,285236
0,8	0,288145	0,291030	0,293892	0,296731	0,299546	0,302338	0,305106	0,307850	0,310570	0,313267
0,9	0,315940	0,318589	0,321214	0,323814	0,326391	0,328944	0,331472	0,333977	0,336457	0,338913
1	0,341345	0,343752	0,346136	0,348495	0,350830	0,353141	0,355428	0,357690	0,359929	0,362143
1,1	0,364334	0,366500	0,368643	0,370762	0,372857	0,374928	0,376976	0,378999	0,381000	0,382977
1,2	0,384930	0,386860	0,388767	0,390651	0,392512	0,394350	0,396165	0,397958	0,399727	0,401475
1,3	0,403199	0,404902	0,406582	0,408241	0,409877	0,411492	0,413085	0,414656	0,416207	0,417736
1,4	0,419243	0,420730	0,422196	0,423641	0,425066	0,426471	0,427855	0,429219	0,430563	0,431888
1,5	0,433193	0,434478	0,435744	0,436992	0,438220	0,439429	0,440620	0,441792	0,442947	0,444083
1,6	0,445201	0,446301	0,447384	0,448449	0,449497	0,450529	0,451543	0,452540	0,453521	0,454486
1,7	0,455435	0,456367	0,457284	0,458185	0,459071	0,459941	0,460796	0,461636	0,462462	0,463273
1,8	0,464070	0,464852	0,465621	0,466375	0,467116	0,467843	0,468557	0,469258	0,469946	0,470621
1,9	0,471284	0,471933	0,472571	0,473197	0,473810	0,474412	0,475002	0,475581	0,476148	0,476705
2	0,477250	0,477784	0,478308	0,478822	0,479325	0,479818	0,480301	0,480774	0,481237	0,481691
2,1	0,482136	0,482571	0,482997	0,483414	0,483823	0,484222	0,484614	0,484997	0,485371	0,485738
2,2	0,486097	0,486447	0,486791	0,487126	0,487455	0,487776	0,488089	0,488396	0,488696	0,488989
2,3	0,489276	0,489556	0,489830	0,490097	0,490358	0,490613	0,490863	0,491106	0,491344	0,491576
2,4	0,491802	0,492024	0,492240	0,492451	0,492656	0,492857	0,493053	0,493244	0,493431	0,493613
2,5	0,493790	0,493963	0,494132	0,494297	0,494457	0,494614	0,494766	0,494915	0,495060	0,495201
2,6	0,495339	0,495473	0,495603	0,495731	0,495855	0,495975	0,496093	0,496207	0,496319	0,496427
2,7	0,496533	0,496636	0,496736	0,496833	0,496928	0,497020	0,497110	0,497197	0,497282	0,497365
2,8	0,497445	0,497523	0,497599	0,497673	0,497744	0,497814	0,497882	0,497948	0,498012	0,498074
2,9	0,498134	0,498193	0,498250	0,498305	0,498359	0,498411	0,498462	0,498511	0,498559	0,498605
3	0,498650	0,498694	0,498736	0,498777	0,498817	0,498856	0,498893	0,498930	0,498965	0,498999
3,1	0,499032	0,499064	0,499096	0,499126	0,499155	0,499184	0,499211	0,499238	0,499264	0,499289
3,2	0,499313	0,499336	0,499359	0,499381	0,499402	0,499423	0,499443	0,499462	0,499481	0,499499
3,3	0,499517	0,499533	0,499550	0,499566	0,499581	0,499596	0,499610	0,499624	0,499638	0,499650
3,4	0,499663	0,499675	0,499687	0,499698	0,499709	0,499720	0,499730	0,499740	0,499749	0,499758
3,5	0,499767	0,499776	0,499784	0,499792	0,499800	0,499807	0,499815	0,499821	0,499828	0,499835
3,6	0,499841	0,499847	0,499853	0,499858	0,499864	0,499869	0,499874	0,499879	0,499883	0,499888
3,7	0,499892	0,499896	0,499900	0,499904	0,499908	0,499912	0,499915	0,499918	0,499922	0,499925
3,8	0,499928	0,499930	0,499933	0,499936	0,499938	0,499941	0,499943	0,499946	0,499948	0,499950
3,9	0,499952	0,499954	0,499956	0,499958	0,499959	0,499961	0,499963	0,499964	0,499966	0,499967
4	0,499968	0,499970	0,499971	0,499972	0,499973	0,499974	0,499975	0,499976	0,499977	0,499978
4,1	0,499979	0,499980	0,499981	0,499982	0,499983	0,499983	0,499984	0,499985	0,499985	0,499986
4,2	0,499987	0,499987	0,499988	0,499988	0,499989	0,499989	0,499990	0,499990	0,499991	0,499991
4,5	0,4999965992									
5	0,499997129									

Fórmulas de Estadísticas descriptivas²

Presentación de los datos	Modo <i>Mo</i>	Mediana <i>Md</i>	Media \bar{X}	Varianza S^2	Desviación Estándar <i>S</i>
Sin agrupar	Valor más frecuente	Nro. de orden: $(n+1)/2$	$\bar{X} = \frac{\sum x}{n}$	$S^2 = \frac{\sum (X - \bar{X})^2}{n-1}$	$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$
Distribución de Frecuencias Simple	Valor más frecuente	Nro. de orden: $(n+1)/2$	$\bar{X} = \frac{\sum (x.f)}{n}$	$S^2 = \frac{\sum (X - \bar{X})^2 . f}{n-1}$	$S = \sqrt{\frac{\sum (X - \bar{X})^2 . f}{n-1}}$
Distribución de Frecuencias Agrupada en Intervalos	Punto medio del intervalo que tiene la mayor frecuencia.	$LIR + \frac{(n+1)/2 - faa}{f} i$	$\bar{X} = \frac{\sum (PM.f)}{n}$	$S^2 = \frac{\sum (PM - \bar{X})^2 . f}{n-1}$	$S = \sqrt{\frac{\sum (PM - \bar{X})^2 . f}{n-1}}$

² Elaborado por María Alejandra Otamendi y Teresa Almendros