

## Práctico 5. Distribuciones bidimensionales

### Tablas de contingencia

1. En una encuesta realizada a un grupo de 200 estudiantes secundarios se les interrogó acerca de su preferencia por estudios literarios, científicos o técnicos, obteniéndose la siguiente información.

Estudios	Hombres	Mujeres
Literarios	60	60
Científicos	42	15
Técnicos	18	5

- Calcule las distribuciones marginales.
- ¿Qué porcentaje de estudiantes prefiere estudio técnico?
- ¿Qué porcentaje de los estudiantes es hombre?
- ¿Qué porcentaje de estudiantes es hombre y además además prefiere estudios técnicos?
- ¿Qué porcentaje de los hombres prefiere estudio técnico?
- ¿Qué porcentaje de los que prefieren dibujo técnico es hombre?

2. A partir del siguiente cuadro sobre preferencia de tipos de comida según edad, contestar las preguntas...

	china	italiana	mexicana	
Menos de 12	9	8	5	22
Entre 12 y 18	13	5	9	27
Más de 18	14	7	15	36
	36	20	29	85

- ¿Qué clase de comida tiene mayor aceptación entre los menores de 12, y en qué porcentaje?
- Si analizamos por clase de comida, ¿cuál es la más preferida por los menores de 12 que por los otros dos grupos de edad?
- De los que prefieren comida china o italiana ¿cuántos tienen más de 18 años?
- ¿Cuántos prefieren comida italiana y además tienen más de 12 años?

3. En la escuela secundaria *Marcos Sastre*, de la ciudad de Buenos Aires, se seleccionó al azar a 350 alumnos para preguntarles por sus materias predilectas.

- ¿Cuál es la población en estudio? ¿Se podría decir que la conclusión que se saque de este estudio se puede extender a todo el alumnado secundario de la ciudad de Buenos Aires? Justifique.
- Complete la tabla de frecuencias a partir de la siguiente información:
  - El 20% del total de los encuestados son hombres y prefieren química.
  - El 24% del total, sin distinción de sexo, prefiere física.
  - Si nos restringimos a los que solo estudian matemáticas, un tercio de ellos son mujeres.

	Matemática	Física	Química	Total
Hombres		75		
Mujeres				
Total	45			350

4. Considere la siguiente distribución de las variables “medios de transporte utilizado” y “cantidad de desplazamientos diarios por los alumnos de una escuela”.

Cantidad de desplazamientos de los alumnos							
Transporte	1	2	3	4	5	6	Total
Subte	1	7	5	11	0	0	24
Colectivo	1	8	1	5	0	1	16
Moto	0	0	1	0	1	0	2
Auto	0	2	1	11	4	0	18
Total:	2	17	8	27	5	1	60

- ¿Qué porcentaje utiliza el colectivo?
- De alumnos que utilizan el subterráneo ¿Qué porcentaje realiza entre 1 y 3 viajes inclusive?
- De los alumnos que realizaron 3 viajes ¿Qué porcentaje utilizan cualquier medio excepto el subterráneo?
- ¿Qué porcentaje de alumnos utiliza el transporte privado (o sea, moto o auto)?
- Si nos restringimos a los que hacen 4 o más desplazamientos ¿qué porcentaje utiliza moto o auto?

5. Los siguientes son los resultados del personal docente según categorías agrupadas y sexo, por grupos de edad en la Universidad de Buenos Aires en el año 2000.



## CENSO DE DOCENTES 2000

Capítulo I - Cuadro 1: Personal docente según categorías agrupadas y sexo, por grupos de edad

Edad	Profesores			Auxiliares			Total		
	Varones	Mujeres	Total	Varones	Mujeres	Total	Varones	Mujeres	Total
Hasta 24	3	9	12	440	611	1051	443	620	1063
25 - 29	59	63	122	1381	1902	3283	1440	1965	3405
30 - 34	194	169	363	1291	1642	2933	1485	1811	3296
35 - 39	364	250	614	1226	1331	2557	1590	1581	3171
40 - 44	480	357	837	901	1156	2057	1381	1513	2894
45 - 49	632	482	1114	608	857	1465	1240	1339	2579
50 - 54	658	464	1122	396	550	946	1054	1014	2068
55 - 59	766	475	1241	246	328	574	1012	803	1815
60 - 64	584	278	862	162	137	299	746	415	1161
65 y más	429	126	555	41	36	77	470	162	632
Total	4169	2673	6842	6692	8550	15242	10861	11223	22084

- a) ¿Qué porcentaje de docentes de la universidad son varones?
- b) ¿Qué porcentaje de los varones son docentes auxiliares?
- c) ¿Qué porcentaje de los docentes auxiliares son varones?
- d) De los mayores de 54 años ¿Qué porcentaje son profesores?
- e) De los auxiliares ¿Qué porcentaje son menores de 30 años?

**Relación entre dos variables cuantitativas, correlación y regresión.**

6. Se administró a 10 sujetos una prueba de dibujo técnico y otra de imaginación creadora obteniéndose:

<b>Imaginación creadora (X)</b>	11	14	11	7	9	11	3	7	6	1
<b>Dibujo (Y)</b>	13	12	10	10	8	6	6	5	3	2

- a) Realizar el diagrama de dispersión.
- b) Sabiendo que:

$$\sum x = 80 \quad \sum y = 75 \quad \sigma_x = 3,8 \quad \sigma_y = 3,53$$

Calcule:  $\bar{x}$  ;  $\bar{y}$  ;  $\sum xy$  ; la covarianza  $cov_{xy}$  , el coeficiente de correlación lineal  $\rho$

7. El número de horas dedicadas al estudio de una asignatura y la calificación obtenida en el examen correspondiente, de ocho personas es:

$x =$ horas de estudio	20	16	34	23	27	32	18	22
$y =$ calificación	6,5	6	8,5	7	9	9,5	7,5	8

$$\sigma_x = 6,06 \quad \sigma_y = 1,15$$

$$\sum x = 192 \quad \sum y = 62 \quad \sum xy = 1534$$

Calcule la covarianza y el coeficiente de correlación.

8. Un hotel dispone de habitaciones similares, pero a diferentes precios en dólares. La demanda observada de sus habitaciones en el último verano ha sido la siguiente:

<b>Precio (x)</b>	100	85	75	50	40
<b>Demanda (y)</b>	25	80	130	150	170

Algunos datos ya calculados:

$x$	$y$	$x \cdot y$
100	25	2.500
85	80	6.800
75	130	9.750
50	150	7.500
40	170	6.800

a) Hallar la correlación sabiendo que

$$\bar{x} = 70 \quad \bar{y} = 111 \quad \sigma_x = 22,14 \quad \sigma_y = 52,38$$

b) ¿Considera que la correlación es buena, por qué?

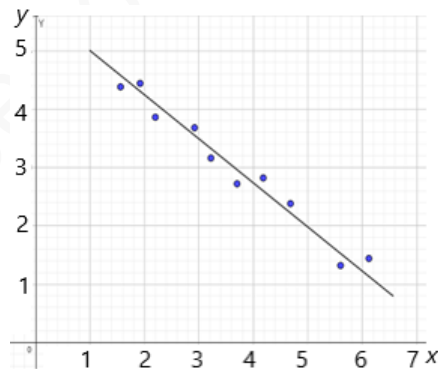
9. Para 4 individuos distintos se midieron las variables “x” e “y” con la finalidad de averiguar si había buena correlación entre ellas y la recta de regresión que las asociaba. Los datos de las variables y algunas cuentas arrojaron estos datos:

$x$	$y$	$x \cdot y$
1	35	35
2	25	50
3	20	60
4	10	40

$$\sigma_x = 1,12 \quad \sigma_y = 9,01$$

Calcule la correlación entre las variables “x” e “y”. ¿Juzga que están muy correlacionadas?

10. Para dos variables X e Y se realizó la nube de puntos y se dibujó la recta de regresión:



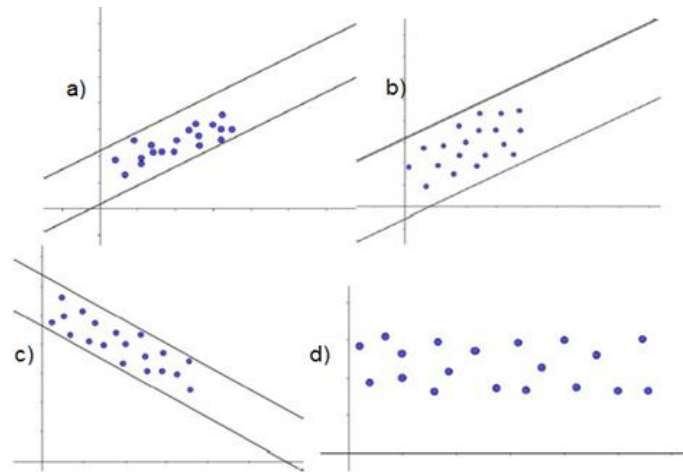
Si le dijese que uno de estos cinco valores:

$$-4,81 ; -0,98 ; -0,12 ; 0,08 ; 0,97$$

es el verdadero índice de correlación que representa a este ejemplo ¿Cuál de los cinco cree que es? ¿Por qué?

11. Para cada uno de los 4 conjuntos de nubes de puntos se calcularon los coeficientes de correlación lineal ( $\rho$ ), y aunque no sabemos qué valor se corresponde con cada gráfico, sí sabemos que valores se obtuvieron. ¿Qué gráfico asignaría a cada uno de los coeficientes?

Correlación:	<b>-0,59</b>	<b>0,09</b>	<b>0,63</b>	<b>0,92</b>
Gráfico letra:				



**Respuestas Práctico 5. Distribuciones bidimensionales**

1.a)

Estudios	Sexo		Totales
	H	M	
L	60	60	120
C	42	15	57
T	18	5	23
<b>Totales</b>	<b>120</b>	<b>80</b>	<b>200</b>

b)  $\frac{23}{200} 100 = 11,5\%$

c)  $\frac{120}{200} 100 = 60\%$

d)  $\frac{18}{200} 100 = 9\%$

e)  $\frac{18}{120} 100 = 15\%$

f)  $\frac{18}{23} 100 = 78,26\%$

2.

a. La **china**, que es preferida por el  $\frac{2}{22} 100 = 9,09\%$

b. La italiana es preferida por el  $\frac{8}{20} 100 = 40\%$  de los menores de 12, contra 25% para la china y 17,24% para la mexicana.

c.  $\frac{14+7}{36+20} 100 = \frac{21}{56} 100 = 37,5\%$

d.  $5 + 7 = 12$

3. a) La población está formada por los alumnos de la escuela de Marcos Sastre. No se puede generalizar al resto de las escuelas porque los alumnos se eligieron al azar **solo en la escuela de Marcos Sastre**, no fue una muestra representativa de la ciudad.

b) Si 20% son hombres y prefieren química, entonces como 20% de 350 es 70 obtenemos la casilla de la fila Hombres en la columna Química. Como condicionado a matemáticas un tercio son mujeres, entonces de los 45 que prefieren matemática, 15 son mujeres. El total de física se



obtiene del 24% de 350, o sea, 84. Las demás frecuencias conjuntas salen a partir de diferencias o sumas, hasta completar:

	Matemática	Física	Química	Total
Hombres	30	75	70	175
Mujeres	15	9	151	175
Total	45	84	221	350

4.

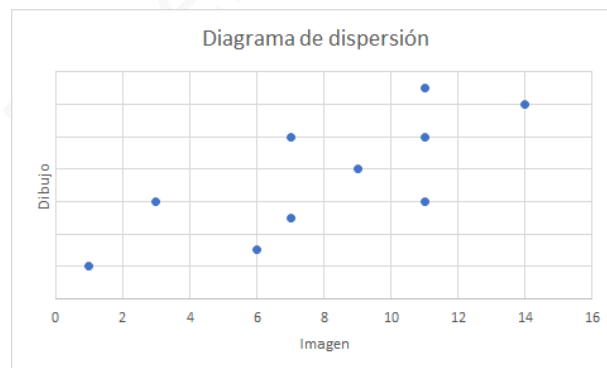
- a)  $16 \cdot 100 / 60 = 26,66 \%$
- b)  $13 \cdot 100 / 24 = 54,17 \%$
- c)  $3 \cdot 100 / 8 = 37,5 \%$
- d)  $20 \cdot 100 / 60 = 33,33 \%$
- e)  $16 \cdot 100 / 33 = 48,48 \%$

5.

- a)  $\frac{10861}{22084} = 49,18 \%$
- b)  $\frac{6692}{10861} = 61,61 \%$
- c)  $\frac{6692}{15242} = 43,90 \%$
- d)  $\frac{1241+862+555}{1815+1161+632} = \frac{2658}{3608} = 73,67 \%$
- e)  $\frac{1051+3283}{15242} = \frac{4334}{15242} = 28,43 \%$

6.

a)



b)  $\bar{x} = 8 \quad \bar{y} = 7,5 \quad \sum x \cdot y = 702 \quad cov_{xy} = \frac{702}{10} - (8 \cdot 7,5) = 10,2$ ,

$$\rho = \frac{cov_{xy}}{\sigma_x \sigma_y} = \frac{10,2}{3,79 \cdot 3,53} = 0,76$$

$$r = \rho^2 \cdot 100 = 57,6 \%$$

Nos dice que solo el 57,6% de la nota en imaginación creadora queda explicada por la nota del dibujo técnico.

7.  $cov_{xy} = 5,75$  ;  $\rho = 0,825$  (Nos indica que es muy bueno por estar cerca del valor 1);

8.  $cov_{xy} = -1100$   $\rho = -0,95$   $r = 90,25 \%$

Que la correlación de  $-0,95$  indica que estamos ante una correlación negativa (a mayor precio, menos demanda) y al ser muy cercana al  $-1$  indica que **es muy buena**.

9. correlación  $\rho = -0,992$  (covarianza  $-10$ ), o sea que es negativa (lo que quiere decir que, a mayor valor de  $x$ , menor valor de  $y$ ).

Como la correlación está cerca del valor  $-1$ , están muy correlacionadas.

10.

$-0,98$  (los valores positivos los descartamos pues la recta tiene pendiente negativa, el valor  $-0,01$  sería en el caso de que casi no haya correlación, sin embargo, es evidente en el gráfico que este no es el caso, el valor  $-4,8$  es imposible porque la correlación es un parámetro que jamás puede ser menor que  $-1$  ni mayor que  $1$ ).

11.

Correlación $\rho$ :	<b>- 0,59</b>	<b>0,09</b>	<b>0,63</b>	<b>0,92</b>
Gráfico letra:	<b>c</b>	<b>d</b>	<b>b</b>	<b>a</b>

En el gráfico d) la correlación es casi nula, por lo tanto, le corresponde  $\rho = 0,09$ . La única relación en que se espera una recta de regresión negativa es en la c), por lo tanto, le asignamos  $\rho = -0,59$ . De las dos que quedan, la a) tiene mejor correlación que la b), por lo tanto, será más cercana a 1: por lo que al a) le asignamos  $\rho = 0,92$  y al b) el  $0,63$ .