

HOJA 1

1. Dada la siguiente distribución de frecuencias:

x_i	n_i	f_i	N_i
2	10	0.25	10
3			
4	15	0.375	30
5		0.25	

Determinar la media, moda, mediana y primer cuartil.

2. Dada la siguiente muestra de estatura de 24 personas:

1.62 1.75 1.60 1.41 1.93 2.00 1.71 1.68 1.60 1.67 1.85
 1.83 1.57 1.54 1.62 1.93 1.84 2.01 1.70 1.85 2.05 1.66
 1.90 1.65

Determinar la media, mediana, varianza y desviación típica para los datos originales y para los datos agrupados.

3. Las calificaciones de la asignatura de Estadística del curso 10/11 se distribuyen de acuerdo a la siguiente tabla para los alumnos presentados en junio,

Calificación	Valor	Alumnos
Suspenso	0	110
Aprobado	1	90
Notable	2	23
Sobresaliente	3	12
Matrícula de Honor	4	2

- a) Represente el polígono de frecuencias.
 b) Calcule la media aritmética, la moda y la mediana.
 c) Calcule la varianza y el cuartil $Q_{3/4}$
4. La siguiente tabla de frecuencias recoge las puntuaciones obtenidas por 100 empleados sanitarios en una encuesta sobre la evaluación de la calidad del trabajo:

PUNTOS	n_i	N_i	f_i	F_i
2 ó menos		6		
3	10			
4			0,14	
5				0,48
6		64		
7		78		
8	10			
9				0,96
10				

Se pide:

- a) Completar la tabla.
 b) Calcular el valor mínimo, los tres cuartiles y el valor máximo.
 c) Calcular la desviación típica y el coeficiente de variación de las puntuaciones.

5. Para obtener información sobre el porcentaje de una sustancia contenida en un preparado químico se analizaron 40 unidades de ese preparado, con los siguientes resultados:

70.2	70.7	72.8	68.7	71.9	64.4	62.4	60.4	62.8	67.0
62.9	65.9	67.5	66,6	67.8	73.5	68.4	63.1	72.3	73.4
70.1	70.0	62.3	65.0	65.2	72.1	66.1	65.5	62.0	63.3
72.4	69.4	66.3	71.3	70.2	68.5	69.1	66.4	65.2	70.1

Suponiendo que un error del 0.5 en el porcentaje del contenido de la sustancia en estudio no es importante, agrupar los datos en clases o intervalos de amplitud 2 y tabular los resultados obtenidos.

6. Los siguientes datos corresponden al número de individuos aquejados de gripe en una ciudad en varios días consecutivos:

12213; 13455; 15066; 12687; 14268; 18307; 12862; 14635; 12344; 12397; 14302; 17068; 13604

- Determinar el número medio de enfermos de gripe.
 - El responsable del almacén que suministra los tratamientos a las farmacias de la ciudad tiene que decidir el número de tratamientos que tiene que tener disponibles para el día siguiente; decide almacenar el número mínimo que permitiese tratar a los enfermos al menos el 90% de los días anteriores. ¿Cuántas dosis tiene que almacenar?
7. Calcular la media aritmética de los siguientes valores, agrupándolos primero por intervalos de amplitud igual a 5 y después por intervalos de amplitud igual a 10:

49	48	43	42	49	41	42	43	43	44	44	51
53	54	51	59	58	57	56	54	51	54	53	64
62	64	63	62	61	62	68	58	67	66	69	

8. El número de pacientes que llegan a urgencias en una noche aumenta al incrementarse la temperatura. Para determinar si la afirmación es cierta, se ha observado la temperatura máxima y el número de pacientes que llegan a urgencias en el turno de noche en los últimos 5 días:

DÍAS	PACIENTES	TEMPERATURA
1	20	39
2	22	40
3	15	30
4	12	30
5	17	32

Se pide:

- Justificar si se acepta la hipótesis, calculando e interpretando el indicador o la medida adecuada.
- Ajustar una recta de regresión que permita estimar el número de pacientes que podrían llegar a urgencias en función de la temperatura máxima diaria.

c) ¿Cuántos pacientes llegarían a urgencias en la noche de un día cuya temperatura máxima ha sido de 35 grados?

9. Se han medido pesos y alturas de seis personas obteniéndose los datos siguientes:

Pesos	65	60	65	63	68	68
Alturas	1.70	1.50	1.68	1.70	1.75	1.80

Se quiere saber:

- ¿Qué medidas están más dispersas los pesos o las alturas?
- ¿Cuál es el coeficiente de variación en cada caso? Conclusión.

10. Se tiene la siguiente tabla de frecuencias relativas correspondiente a una variable bidimensional (x; y):

x\y	10	15	20	25	30	35
8	0.04	0.05	0.05	0.03	0.00	0.05
10	0.06	0.10	0.00	0.07	0.05	0.10
12	0.12	0.05	0.05	0.03	0.10	0.05

Teniendo en cuenta que el tamaño de la muestra es 200, se pide:

- Número de elementos que tienen $y < 22$ y $x > 9$.
- Distribuciones marginales.
- Distribución condicionada de x para $y = 25$.
- Distribución condicionada de y para $x = 12$.
- Centro de gravedad (\bar{x}, \bar{y})
- Tercer cuartil de la y.

11. Una explotación agrícola ha tenido en los siete últimos años unos gastos de producción X y unos ingresos totales Y que son, en millones de euros, los indicados en la siguiente tabla:

X	1.9	2.2	2.9	3.6	3.8	4.6	5.5
Y	5.5	7.4	9.8	11.6	11.6	12.2	11.2

- Hacer una representación gráfica de los datos.
- Obtener la recta que mejor explica los ingresos totales en función de los gastos de producción.
- ¿Cuáles son los ingresos totales que deben esperarse para un año en el que los gastos de producción fueron de 4 millones de euros?

12. La concentración, X e Y, de dos sustancias en la sangre parece estar relacionada. Para estudiar esta posible relación, se miden estas cantidades en 30 personas, obteniéndose los siguientes resultados:

$$\sum x_i = 41,2; \sum y_i = 63,8; \sum y_i x_i = 118,7; \sum x_i^2 = 188,2; \sum y_i^2 = 296,4$$

Hallar la recta de regresión de Y sobre X y el coeficiente de correlación lineal.

13. Se piensa que la temperatura de un cierto fluido puede inferirse a partir de la concentración de un cierto isótopo de oxígeno en ese

fluido. Dada la siguiente tabla de valores medidos de ambos parámetros.

$\delta^{18}O(\text{‰})$	Temperatura $^{\circ}\text{C}$
-10	5
-15	-10
5	0
10	10
-8	-15
-7	-10
5	10
20	20
10	15
0	5

- Determinar la recta de regresión correspondiente y representar gráficamente.
- Determinar el coeficiente de correlación lineal y determinar qué proporción de la variabilidad de los datos es explicada por el ajuste propuesto.

14. La siguiente tabla muestra el nivel de colesterol (X) y la edad (Y) de un grupo de varones:

X/Y	(160, 200)	(200, 240)	(240, 280)
[20,40)	25	13	2
[40,60)	20	20	10
[60,80)	15	20	15

- Calcular la covarianza e interpretar su valor.
 - Si se considera que el colesterol es alto a partir de 240, ¿qué porcentaje de varones entre 40 y 60 años tiene colesterol alto?
15. Se desea investigar si existe relación entre X: el número de ficheros que se mandan a imprimir en un estanco e Y: el tiempo, en horas, que tarda la impresión (Y). Después de haber registrado la información de 43 estancos, se obtiene la siguiente tabla de doble entrada:

X \ Y	[0.5; 1.5)	[1.5; 2.5)	[2.5; 3.5)	[3.5; 4.5)	[4.5; 5.5)	$n_{i\cdot}$
5	4	1	0	0	0	5
10	1	5	1	0	0	7
15	0	3	8	2	0	13
20	0	1	2	6	0	9
25	0	0	2	3	2	7
30	0	0	0	1	1	2
$n_{\cdot j}$	5	10	13	12	3	43

Se pide:

- Calcula el percentil 80 del número de ficheros impresos. ¿Qué porcentaje de estancos han imprimido con tiempo de duración entre 2

y 4 horas?

- b) Calcula el coeficiente de variación de cada una de las dos variables X e Y e interpreta los resultados.
- c) Calcula el coeficiente de correlación lineal entre X e Y.

16. Se efectúa una encuesta sobre el hábito de fumar y el ejercicio físico. Se considera que una persona no realiza ejercicio físico si hace menos de 1 hora de deporte a la semana, realiza poco ejercicio si practica deporte entre 1 y 4 horas y mucho ejercicio si dedica entre 4 y 6 horas a la semana. Se sabe:

- La frecuencia relativa de fumadores que no realizan ejercicio físico es 0.25,
- la frecuencia relativa de no fumadores que realizan poco ejercicio físico es 0.2,
- la frecuencia relativa marginal de ser fumador es 0.4,
- la frecuencia relativa marginal de no realizar ejercicio físico es 0.5,
- la frecuencia relativa de fumadores que hacen mucho ejercicio es de 0.05.

Se pide:

- a) Si se sabe que 30 encuestados son no fumadores y realizan mucho ejercicio físico, ¿cuántas personas fueron encuestadas?
- b) Construye una tabla de doble entrada con las frecuencias absolutas.
- c) Calcula la media aritmética y la mediana del tiempo dedicado para hacer deporte.
- d) ¿Cuántas horas de deporte a la semana tiene que hacer una persona para pertenecer al 15% que más ejercicio hace? Calcula el coeficiente de variación en el tiempo dedicado a hacer deporte.

17. Una compañía discográfica ha recopilado la siguiente información sobre 20 grupos musicales, a saber, el número de conciertos dados este verano y las ventas de discos de estos grupos (en miles de CDs), obteniendo los siguientes datos:

	Conciertos		
CDs	20	40	60
1-6	3	2	1
6-11	1	4	1
11-16	2	1	5

- a) ¿Cuántos CDs han vendido el 15% de los grupos que más venden?
- b) ¿Cuál es el número de conciertos más habitual que han dado esos grupos?
- c) Si un grupo musical ha vendido 1800 CDs, ¿podríamos prever el número de conciertos que darán este verano? Justifica la respuesta y en caso de ser afirmativa calcúlalo.

18. Las estaturas y pesos de 10 jugadores de baloncesto de un equipo son:

Altura	186	189	190	192	193	193	198	201	203	205
Peso	85	85	86	90	87	91	93	103	100	101

- a) ¿Podríamos afirmar que a más altura más peso? Justifique la respuesta.
- b) ¿Podríamos predecir el peso de un jugador en función de su altura? Justifique la respuesta. ¿Cuál es el peso estimado de un jugador que mide 202 cm?
- c) ¿A partir de qué altura se encuentra el 27% de los más altos?
- d) ¿Entre qué valores se encuentra el 30% de los pesos centrales?
19. Los estudiantes en situación de examen desarrollan una gran diversidad de reacciones ansiosas que pueden ir desde un control conductual adecuado (en ocasiones sólo es aparente), hasta un estado de descontrol emocional intenso, que se le hace inmanejable y que llega a bloquear su capacidad para expresar sus conocimientos; esta última situación se conoce habitualmente como inhibición de examen. Se estudió la nota media de un grupo de estudiantes en una asignatura con respecto a su nivel de ansiedad ante los exámenes que se categorizó en nivel de ansiedad bajo, medio y alto. Las calificaciones se han agrupado en 3 intervalos suponiendo uniformidad dentro de los mismos. Todos los datos se recogieron en la siguiente tabla:

Ansiedad	1-4	4-7	7-10
Bajo	1	4	2
Medio	2	5	1
Alto	6	0	0

- a) ¿Qué nivel de ansiedad tienen como mucho el 30% de los estudiantes con más ansiedad?
- b) Haz una representación gráfica en la que aparezca el nivel de ansiedad y la calificación media.
- c) ¿Entre qué calificaciones se encuentran el 40% de las calificaciones centrales de los estudiantes?
20. En un centro de salud hacen el seguimiento de la presión arterial a sus pacientes, y tras analizar los resultados los dietistas afirman que aquellos que tienen sobrepeso tienen una presión arterial superior. Para simplificar los cálculos hemos tomado una muestra representativa de 10 individuos. A partir de los siguientes datos:
- X =Peso(kg) 72 76 78 81 89 95 108 115 120 130
- Y =Presión(mmHg) 115 121 125 130 141 150 165 171 177 178
- a) ¿Es cierta la afirmación de los dietistas? Justificar la respuesta.
- b) ¿Podríamos predecir el peso de un paciente a partir de su presión arterial? Justificar la respuesta.
- c) Si la respuesta del ítem anterior es Si, predecir el peso de un paciente con una presión arterial de 160.

21. Los datos de la producción de un medicamento a nivel mundial en millones de dosis y el precio por dosis en euros al año siguiente en los últimos años fueron:

Años	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Producción	20	18	22	15	15	15	12	14	25	30
Precio	15	20	17	30	22	30	40	35	20	15

- a) Ajusta la recta de regresión que explique el precio de la dosis en función de la producción.
- b) Si en el año 2022 se han producido 16 millones de dosis, ¿qué precio se espera para el año 2023?
- c) ¿Puede considerarse que el modelo lineal es adecuado? Justifica la respuesta.
22. Se han analizado los tiempos de espera (en minutos) en un servicio técnico:
2; 5; 10; 12; 4; 4; 5; 17; 11; 8; 9; 8; 12; 21; 6; 8; 7; 13; 18; 3;
- Con las clases $(0,4]$, $(4,8]$, etc.
- a) Construir la tabla de frecuencias asociada a la variable de interés y representarla gráficamente.
- b) Calcular una medida de dispersión de los datos.
- c) Calcular el intervalo de tiempo de espera del 50% de la distribución central de los datos.