

Estadística Descriptiva II

Distribuciones bidimensionales de frecuencias

TEMA 3

Problema 1

1.- El director de un establecimiento comercial quiere saber si el aumento en el número de clientes potenciales que entran en sus almacenes se traduce en un aumento de sus ventas (cientos de euros). Para ello observa ambas variables durante una semana, de lunes a sábado, y los datos son:

L	M	X	J	V	S
87	63	70	55	90	105
120	85	90	63	110	150

- a) Calcular el número medio de clientes potenciales y las ventas medias diarias
- b) Analizar la interdependencia entre ambas variables. ¿las variables X e Y son estadísticamente independientes?
- c) Predecir las ventas que tendrá un día en que entran 130 personas en el establecimiento y dar una medida de la fiabilidad de dicha predicción.
- d) Calcular la varianza residual e interpretar su significado

Problema 2

En una empresa de transportes trabajan 7 conductores. Los años de antigüedad de permisos de conducir y el número de infracciones cometidas en el último año por cada uno de ellos son los siguientes:

Antigüedad (Años)	Número infracciones
1	2
2	3
3	3
2	5
5	3
6	0
3	1

- Predecir el número de infracciones de un conductor que tiene 4 años de antigüedad
- Y si un conductor ha cometido 4 infracciones en el último año
¿Cuántos años de antigüedad de permiso podemos decir que tiene?
- ¿Cómo evoluciona el nº de infracciones a medida que aumenta la antigüedad de los conductores?
- Cuál es la variabilidad en el nº de infracciones que no queda explicada por la antigüedad de los conductores?
- Se plantea bonificar un 10% de la prima del seguro al 15% de conductores con mayor antigüedad. ¿Qué antigüedad mínima ha de tener un conductor para tener acceso a la bonificación?
- Al 5% de conductores con mayor nº de infracciones se les va a penalizar con un 10% de recargo en la prima del seguro. ¿Hasta qué número de infracciones un conductor queda libre de dicho recargo?
- ¿Qué distribución es más homogénea, la de la antigüedad o la de las infracciones?
- ¿Qué puedes decir de la interdependencia entre las dos variables?

Problema 3

Se trató de ajustar un modelo de regresión lineal simple entre x e y

para analizar en una compañía a lo largo de un año la relación entre las variables

X: coste salarial mensual (en 10^4 euros)

Y: indicador de productividad (en puntos)

Y se dispone de los siguientes cálculos:

$$n=12 \quad \bar{x} = 17 \quad \bar{y} = 12$$

$$\sum_{i=1}^n x_i^2 = 5732 \quad \sum_{i=1}^n y_i^2 = 3593 \quad \sum_{i=1}^n x_i y_i = 4135$$

- ¿En qué medida dirías que el coste salarial explica mediante una relación lineal el indicador de productividad?
- Predecir la productividad un mes en que el coste salarial asciende a 200.000 euros y dar una medida de la fiabilidad de dicha predicción
- ¿Cómo puedes comparar la dispersión de las variables “coste salarial” y “productividad”?
- Si hemos de recortar la masa salarial en 3.000 euros ¿Cómo dirías que se verá afectada la productividad de los empleados? Razónalo

Problema 4

En una empresa se pretende analizar la relación existente entre: las ventas totales (X) y las exportaciones (Y), expresadas en Millones de euros.

Disponemos de los siguientes datos:

- ✓ Las dos rectas de regresión se cortan en el punto (6,4).
- ✓ Por cada millón de euros que se incrementan las exportaciones, el incremento que experimentan las ventas totales, bajo el modelo lineal de Y sobre X, es de 1,15 M euros.
- ✓ Si se incrementan en un millón de euros las ventas totales, las exportaciones aumentarían en 810.000 euros.
- ✓ La dispersión relativa de las exportaciones (Y), medida en términos de su coeficiente de variación, ha sido del 50 %.

a) Con estos supuestos, predecir las exportaciones de la empresa en un período en que las ventas totales han sido de 25 Millones de euros y dar una medida de la fiabilidad de dicha predicción

b) Calcular el coeficiente de correlación e interpretarlo

c) Comparar la dispersión relativa de ambas distribuciones

d) Calcular e interpretar la varianza residual y la varianza explicada

Problema 5

En una empresa se analiza la posible existencia de una relación lineal entre la edad (X) de sus trabajadores y el número de días de baja a lo largo de un año.

Se han obtenido los siguientes datos correspondientes a 5 trabajadores

$$\bar{x} = 44,2 \quad \bar{y} = 8,4 \quad \sum_{i=1}^n x_i^2 = 10.439 \quad \sum_{i=1}^n y_i^2 = 368 \quad \sum_{i=1}^n x_i y_i = 1.937$$

- a) Obtener una medida de la correlación entre ambas variables e interpretarla
- b) Hallar una predicción del tiempo de baja de un trabajador de 32 años
- c) ¿Cuál es la variabilidad en los períodos de baja que se explica por la edad de los trabajadores? ¿Qué porcentaje representa esa variabilidad?
- d) Calcular la pendiente de la recta de regresión de y sobre x e interpretarla en el contexto del problema
- e) Razonar si la edad media es más o menos representativa que el número medio de días de baja para los trabajadores de dicha empresa
- f) Si nos facilitan además las edades de los 5 trabajadores de la empresa que son:

50 44 37 60 30

Dar una medida que permita analizar la simetría de la edad de esos trabajadores e interpretarla. Interpretar también los valores que utilices en la fórmula correspondiente.

Problema 6

El número de cetano se emplea como indicador de la calidad de ignición del combustible utilizado en un motor diésel.

En un estudio se trató de ajustar mediante un modelo lineal dicho indicador (Y) en función del índice de yodo X (en gramos)

Se analizó una muestra de 14 combustibles y se obtuvieron los siguientes cálculos:

$$\begin{aligned}\sum_{i=1}^n x_i &= 1.307,5 & \sum_{i=1}^n y_i &= 779,2 & \sum_{i=1}^n x_i^2 &= 128.913,93 \\ \sum_{i=1}^n y_i^2 &= 43.745,22 & \sum_{i=1}^n x_i y_i &= 71.347,3\end{aligned}$$

- a) Predecir utilizando un modelo lineal el número de cetano para un combustible que tiene un índice de yodo de 115 y explicar el significado del coeficiente de regresión en ese modelo.
¿Cuál es la fiabilidad de dicha predicción?
- b) Dar una medida de la bondad del ajuste e interpretarla
- c) ¿Cuánta variabilidad en el número de cetano no se explica por el índice de yodo?
Indicar también qué % representa ese valor

Se han obtenido los siguientes datos adicionales sobre el número de cetano de la muestra de 14 combustibles

- i. La mitad de ellos tiene un índice de cetano comprendido entre 87 y 114
 - ii. La mitad de los combustibles de la muestra tienen un índice de cetano menor o igual que 100
 - iii. Los índices de cetano mínimo y máximo obtenidos en la muestra de combustibles han resultado de 51 y 173
 - iv. El índice de cetano más frecuente ha sido de 92.
- d) Construir un box plot para el número de cetano e interpretarlo
 - e) ¿Qué valor medio es más representativo, el del número de cetano o el del índice de yodo? Explicar por qué.

Problema 7

Un artículo especializado analizó cómo la intensidad del enfado en los berrinches de los niños podría estar relacionada con la duración de la rabieta. Se analizó para un grupo de niños un indicador Y (en puntos) medidor de la intensidad (gritar, empujar o tirar objetos) en función de la duración del berrinche medida en minutos (X).

I_i	[0,2)	[2,4)	[4,11)	[11,20)	[20,30)	[30,40)
y_i	136	92	71	26	7	3

- Obtener una medida de la correlación entre ambas variables e interpretarla
- Hallar una predicción del tiempo que duró la rabieta en un niño en el que se obtuvo un indicador de 81 indicando la fiabilidad de dicha predicción
- ¿Cuál es la variabilidad en la intensidad de las rabietas de ese grupo de niños que se explica por el tiempo que duran? ¿Qué porcentaje representa esa variabilidad?
- ¿Cuánto aumenta/disminuye el indicador de intensidad por cada 2 minutos más de berrinche?
- ¿Cuál es el indicador medio de intensidad del enfado? Razonar si es o no un valor representativo de los datos
- Calcular e interpretar en el contexto del problema el primer cuartil de X y el 90 Percentil de Y.

Problema 8

Datos de los billetes vendidos por Renfe durante un mes concreto. Abajo se muestra la tabla de contingencia correspondiente a las variables del estudio:

Variable X: Clase (turista, preferente)

Variable Y: Tipo de tren (alta velocidad, larga distancia, vía estrecha).

Clase /Tipo tren	Vía estrecha	Convencional	Alta velocidad	
Turista	400		800	1800
Preferente	100	400		
	500	1000	1500	

- Completa la tabla de contingencia anterior y conteste a las siguientes preguntas:
- ¿Cómo se llama el valor 800 y qué símbolo le corresponde?
- Proporción de los billetes que son de clase turista. Cálculo, símbolo y nombre estadístico
- Calcula $n_{2.}$ e interpretar su significado
- Proporción de los billetes vendidos en preferente que son de alta velocidad. Cálculo, nombre estadístico
- Proporción de billetes vendidos que son de clase turista y de tipo convencional. Cálculo y nombre estadístico.
- Distribución de frecuencias relativas marginales para la variable clase
- Distribución de frecuencias relativas de la variable clase condicionada a tren de vía estrecha

Problema 9

Disponemos de los datos del número de vehículos que han llegado al taller durante un mes para arreglos de chapa, eléctricos o mecánicos, con indicación de si lo han hecho por la mañana o por la tarde

Arreglo Y→ Entrada X↓	Arreglos de chapa	Arreglos eléctricos	Arreglos mecánicos	$n_{i.}$
Mañana	37	81	100	218
Tarde	123	95	47	265
$n_{.j}$	160	176	147	n=483

- Obtener la distribución marginal de la variable “entrada”:Y
- ¿Cuál es la frecuencia conjunta de “tarde” y “arreglos eléctricos”?
- ¿y cuál es la proporción de clientes que llegan de “tarde” y requieren “arreglos eléctricos”?
- ¿Cuál es el número total de “arreglos de chapa” y qué proporción representan sobre el total de vehículos atendidos este mes en el taller?
- Obtener la frecuencia absoluta marginal de “tarde” e indicar su significado
- Frecuencia relativa marginal de “tarde” y expresar su significado
- Obtener la distribución de frecuencias de arreglos mecánicos condicionada a “mañana”:
- ¿Qué proporción de coches que entraron por la mañana, requirieron de arreglos mecánicos? Un 45,87%
- Obtener el valor de $f_{3.}$ e indicar su significado en el contexto del problema
- Calcular el valor de $n_{2.}$ e indicar qué significa en el contexto del problema
- Proporción de vehículos que han entrado en el taller por la tarde y requieren arreglos de chapa. Calcular el valor e indicar el concepto estadístico correspondiente
- Proporción de vehículos que, habiendo entrado en el taller por la tarde, necesitan reparación de chapa. Calcularla e indicar el concepto estadístico

Problema 10

Disponemos de los siguientes datos del PIB municipal per cápita X (en miles de euros) y la tasa de paro (en %) (Y) correspondientes a 15 municipios

$$\bar{x} = 21,2 \quad \bar{y} = 10,5 \quad \sum_{i=1}^n x_i^2 = 7.033 \quad \sum_{i=1}^n y_i^2 = 2.158 \quad \sum_{i=1}^n x_i y_i = 2.964,12$$

- a) Obtener una medida de la interdependencia entre ambas variables e interpretarla
- b) Obtener una predicción de la tasa de paro de un municipio con un PIB municipal per cápita de 23.300 euros
- c) ¿Cuál es la variabilidad en la tasa de paro que no se explica por el PIB per cápita de los municipios? ¿Qué porcentaje representa esa variabilidad?
- d) Calcular la pendiente de la recta de regresión de y sobre x e interpretarla en el contexto del problema

Si nos facilitan además las tasas de paro de los 15 municipios que son:

18,6 6,7 11,2 9,3 11,4 10,6 11,2 6,7 11,2 9,3 10,7 9,3 11,2 9,3 5,7

- e) Dibujar un boxplot y extraer conclusiones acerca de la tasa de paro en esos municipios.