

# SEMINARIOS

(Problemas de exámenes de años anteriores)

## Estadística. 1º Grado en Informática

### Seminario de Estadística Descriptiva Unidimensional y Bidimensional

1. Se ha realizado un control de calidad en una muestra aleatoria de 200 tubos fluorescentes de un determinado tipo, para determinar su duración en horas de funcionamiento en condiciones prefijadas y se obtuvieron los siguientes resultados:

Duración	Número de tubos
[800,2200]	14
(2200,3800]	88
(3800,5000]	84
(5000,5800]	8
(5800,7200]	6

- a) Si del total de la muestra se elimina el diez por ciento de los tubos con menor duración, determinar el valor mínimo de la duración de los tubos restantes de la muestra.
- b) ¿Cuál es la duración más frecuente de los tubos?
- c) En un control de calidad que se había realizado anteriormente sobre una muestra aleatoria de 300 tubos se obtuvo una duración media de 3000 horas y una varianza de 1300925. Según estos datos, ¿en cuál de las dos muestras la duración de los tubos es más homogénea?
2. Se ha consultado a 50 automovilistas sobre sus ingresos totales anuales y el precio de su coche. Los datos obtenidos fueron los siguientes:

Precio (en miles de euros)	Ingresos ( en miles de euros)			
	[5,15]	(15,25]	(25,50]	(50,100]
[6, 16]	13	9	0	0
(16, 24]	0	19	0	0
(24, 50]	0	0	8	0
(50, 100]	0	0	0	1

- a) ¿Cuál es el precio de coche más frecuente de los automovilistas que tienen unos ingresos anuales entre 15 y 50 mil euros?
- b) ¿Qué ingresos mínimos tienen la mitad de los automovilistas que más ingresos tienen?

- c) ¿Cuál de las dos variables es más homogénea?
  - d) Calcula la recta de regresión que determina el precio del coche en función de los ingresos.
  - e) Determina el grado de relación lineal entre las variables. Comenta el resultado.
  - f) ¿Cuál sería el precio para un automovilista que tuviera unos ingresos de 22 mil euros?
  - g) ¿Qué fiabilidad tiene el resultado obtenido en el apartado anterior?
3. En un estudio se analizó la antigüedad de un modelo de vehículo ( $X$ ) y el nivel de monóxido de carbono ( $Y$ , en gramos/km) emitido por el mismo, obteniéndose los siguientes datos:

Antigüedad	Nivel de monóxido de carbono (gramos/km)		
	[0,50]	(50,120]	(120,300]
[0, 5]	30	4	1
(5, 10]	2	8	3
(10, 15]	0	2	22

- a) Calcula la antigüedad más frecuente de este modelo de vehículos.
- b) ¿Cuál de las dos variables es más homogénea? Justifica la respuesta.
- c) Para los coches que tienen más de 5 años de antigüedad, calcula el nivel de monóxido de carbono máximo que emiten el 10% de los vehículos que menos contaminan.
- d) ¿Existe relación lineal entre las variables? Justifica la respuesta.
- e) Calcula la recta de regresión que determina el nivel de monóxido de carbono que emite el coche en función de la antigüedad del mismo. Calcula el coeficiente de determinación. Interpreta el valor obtenido del coeficiente de determinación.
- f) Compara los resultados obtenidos en el modelo lineal con los modelos proporcionados en las salidas y elige qué modelo se adecua mejor a los datos. Razona la respuesta.
- g) Calcula cuál será el nivel de monóxido de carbono emitido por un coche que tiene 8 años de antigüedad según el modelo elegido en el apartado anterior. ¿Es fiable el dato obtenido?

4. Se pretende estudiar si existe relación entre los ingresos mensuales y el ahorro de un grupo de familias de un barrio conflictivo de una ciudad. Se seleccionan 20 familias al azar y se considera  $Y$ , el ahorro que realizan cada mes (en cientos de euros) y los ingresos mensuales  $X$ , medido en cientos de euros. Los datos aparecen en la tabla adjunta.

$X Y$	$0 - 0.2$	$0.2 - 1$	$1 - 3$	$3 - 7$
$0 - 2$	2	0	0	0
$2 - 8$	0	5	1	0
$8 - 12$	0	0	4	1
$12 - 20$	0	0	2	5

- Obtener una estimación del ahorro de una familia que ingresa 830 euros mensuales.
- Calcula el coeficiente de determinación. Interpreta dicho valor.

### Seminario de Probabilidad y Modelos de Distribuciones

- En una guardería infantil, el 60% de los bebés son niñas. De los niños el 35% son menores de 20 meses. El 20% de las niñas tienen menos de 20 meses. Se selecciona un bebé al azar:
  - Calcular la probabilidad de que el bebé seleccionado sea menor de 20 meses.
  - Si el bebé seleccionado tiene menos de 20 meses, ¿cuál es la probabilidad de que sea una niña?
- La producción de una factoría se realiza en cuatro máquinas  $M_1$ ,  $M_2$ ,  $M_3$ , y  $M_4$ . La producción diaria para  $M_1$  es 600, para  $M_2$  500, para  $M_3$  350 y 250 para  $M_4$ . Además se sabe que los porcentajes de piezas defectuosas producidas por cada una de las máquinas son los siguientes: 4% para  $M_1$ , 3,5% para  $M_2$ , 4,6% para  $M_3$  y 2% para  $M_4$ .
  - Si las piezas se almacenan conjuntamente ¿cuál es la probabilidad de que, al extraer una pieza al azar, esta sea defectuosa?
  - Se ha extraído al azar una pieza que ha resultado ser defectuosa, ¿cuál es la probabilidad de que proceda de  $M_2$ ?
- Un club profesional de atletismo de élite cuenta con un 60% de hombres, un 30% de mujeres y un 10% de atletas infantiles. La probabilidad de que un atleta suba al pódium en una prueba de atletismo si es hombre es de 0.08, de 0.14 si es mujer y de 0.21 si es niño. Obtener:
  - Probabilidad de que, seleccionado un atleta del club al azar, éste suba al pódium.
  - Si un atleta ha subido al pódium, ¿cuál es la probabilidad de que sea mujer?
- La probabilidad de que un niño de 10 años repita curso es de 0.1.
  - si se seleccionan a 12 niños al azar, ¿Cuál es la probabilidad de que 3 o más repitan curso?

- b) Si se seleccionaran al azar 20 niños, ¿Cuál es la probabilidad de que 3 niños repitan curso?
- c) Si se seleccionaran al azar 200 niños, ¿Cuál es la probabilidad de que menos de 18 niños repitan curso?

5.- El volumen que una máquina de llenado automático deposita en latas de una bebida gaseosa tiene una distribución normal con media 34 cl. y desviación típica 1,5 cl.

- a) Si se desechan aquellas latas que tienen menos de 33 cl., ¿cuál es la probabilidad de que una lata sea desechada?
- b) Si se tienen 10 latas llenadas con esta máquina, ¿cuál es la probabilidad de que ninguna lata sea desechada?
- c) Si se tienen 500 latas llenadas con esta máquina, ¿cuál es la probabilidad de que al menos 100 latas sean desechadas?

6.- Una empresa dedicada a fabricar ordenadores compra los chips a una empresa que no es “pirata”. La probabilidad de que un chip de una empresa que no es “pirata” sea defectuoso es del 5%.

- a) Para fabricar un tipo de ordenador utiliza 10 chips, ¿cuál es la probabilidad de que este tipo de ordenador tenga más de 2 chips defectuosos?
- b) Para fabricar un ordenador portátil utiliza 60 chips, ¿cuál es la probabilidad de que un ordenador portátil tenga menos de 10 chips defectuosos?
- c) Para fabricar un dispositivo móvil utiliza otro tipo de chips, que tienen una probabilidad de ser defectuosos del 30%. Si se necesitan 40 chips de este tipo, ¿cuál es la probabilidad de que un dispositivo móvil tenga más del 5% de chips defectuosos?

7.- Supongamos un sistema con 9 componentes que requiere para su funcionamiento que al menos 6 estén disponibles. La probabilidad de funcionamiento de cada una de las componentes, que funcionan de forma independiente, es 0.95.

- a) Calcular la fiabilidad del sistema (probabilidad de que funcione).
- b) Calcular la probabilidad de que estén funcionando entre 3 y 7 componentes, ambos valores incluidos.
- c) Si tenemos un segundo sistema con 40 componentes que requiere para su funcionamiento que al menos 30 estén disponibles, y la probabilidad de funcionamiento de sus componentes es del 0.85, calcular la fiabilidad de este segundo sistema (probabilidad de que funcione).

## Seminario de Inferencia

1. Se quiere comparar la cantidad de CO<sub>2</sub> emitido durante el proceso de fabricación de un producto cuando se utilizan dos métodos diferentes. En una experiencia de laboratorio se han obtenido dos muestras de tamaños 7 y 5. Los datos obtenidos son:

Método 1	123	129	128	116	122	126	131
Método 2	173	182	181	175	179		

- a) Contrastar, con un nivel de significación del 10%, si se puede admitir que los dos métodos dan resultados con la misma varianza.
  - b) Calcular un intervalo de confianza para la diferencia de medias con una confianza del 95%. ¿Se pueden considerar iguales las medias?
2. Se quiere comparar la resistencia al calor de dos materiales diferentes que se utilizan para la fabricación de componentes de una fuente de alimentación. Para hacer el estudio, se recogen muestras del tiempo de exposición a un calor extremo que soportan las piezas fabricadas con ambos materiales. Las variables tiempo de resistencia con los distintos tipos de materiales se suponen independientes y con distribución normal. Contesta razonadamente a las siguientes preguntas:
  - a) ¿En qué muestra hay un tiempo medio de resistencia mayor?
  - b) ¿En qué muestra hay una mayor variabilidad en los resultados?
  - c) ¿Son estas diferencias significativas, o por el contrario puede admitirse que las medias y varianzas son iguales? (utilizar un nivel de significación del 5%)
3. Un instituto de dietética quiere comparar la efectividad de dos dietas. Se selecciona aleatoriamente una muestra de 10 individuos de una población de personas con exceso de peso. A 5 personas se les suministra la dieta A y a los 5 restantes, la dieta B.

Las pérdidas de peso medidas en kg al cabo de mes, son las siguientes:

Dieta A	4,2	3,8	2,8	3	3,5
Dieta B	2,2	2	1,8	2,7	1,9

Suponemos que las variables son normales e independientes:

- a) Estudie mediante un intervalo de confianza, al 95% de confianza, si existen diferencias significativas en la variabilidad de las pérdidas de peso con ambas dietas.
  - b) Teniendo en cuenta el apartado anterior, realice un contraste de hipótesis para decidir si existen diferencias significativas entre las medias de las pérdidas de peso con ambas dietas.
4. Se realiza un estudio para determinar si la combinación de determinados juegos infantiles mejora el aprendizaje de las matemáticas en niños de primaria. Se cree que este nuevo método de aprendizaje (Método A), puede contribuir a la mejora de su atención e interés. Se selecciona aleatoriamente una muestra de 8 niños, a los que se les aplica este nuevo método durante un año. Otra muestra de 8 niños aprende con el método tradicional (Método B). Se establece un sistema de puntuación que evalúa en qué medida se han cumplido los objetivos que se pretenden alcanzar, de forma que una mayor puntuación implica un mejor rendimiento y aprendizaje de los niños. Al finalizar el curso se registran las puntuaciones obtenidas, que se recogen en la siguiente tabla:

Método A	203	229	215	220	223	223	208	228
Método B	221	207	185	203	187	190	195	204

Suponiendo que las variables son normales e independientes, se pide:

- a) Realice un contraste de hipótesis, a un nivel de significación del 5%, para decidir si existen diferencias significativas en la variabilidad de las puntuaciones con ambos métodos.
  - b) Teniendo en cuenta el apartado anterior, realice mediante un intervalo de confianza, a un nivel de confianza del 95%, para estudiar si existen diferencias significativas entre las medias de las puntuaciones obtenidas con estos dos métodos de aprendizaje.
5. Una empresa que proporciona servicios en Internet ha decidido comparar el tiempo de respuesta de dos buscadores que ha diseñado. Suponiendo que el tiempo de respuesta se puede modelizar mediante una distribución normal, se extrae una muestra aleatoria de los tiempos de respuesta en ambos buscadores, resultando:

Buscador 1	1.12	0.91	0.95	1.02	0.98	0.87
Buscador 2	0.72	0.81	0.82	0.90	0.74	0.89

- a) Obtener un intervalo de confianza al 95% para el cociente de varianzas. En base al intervalo calculado, ¿se puede considerar que las varianzas de los tiempos de respuesta pueden ser iguales? Razonar la respuesta.
- b) Contrastar la hipótesis de que el tiempo medio de respuesta del buscador 1 es inferior al tiempo medio de respuesta del buscador 2, con un nivel de significación del 5%.