

RELACIÓN DE EJERCICIOS DE ESTADÍSTICA DESCRIPTIVA.

Ejercicio 1: (Medidas de centralización y dispersión)

En un centro hospitalario de la provincia de Sevilla se ha tratado, con un nuevo medicamento llamado SINDOLORCABEZON, durante 5 días a un grupo de pacientes, todos ellos padecen de jaqueca crónica (se despiertan todos los días con dolor de cabeza). Se realiza un estudio sobre el nº de días que un paciente sufre mejoría con el anterior medicamento obteniendo la tabla:

Valores x_i	Frecuencias n_i
0	100
1	250
2	300
3	500
4	450
5	2000

Realizando el gráfico adecuado y hallando los promedios (Media aritmética, Media armónica, Media geométrica, Moda, y Mediana), indicar cuál sería el que mejor representaría los datos, (Contesta razonadamente y con el mayor detalle posible)

Calcula también el porcentaje de pacientes que sienten mejoría con el medicamento en todos los días del tratamiento.

¿Por qué no calculamos el coeficiente de variación para ver la representatividad de la media? ¿Habría que hallarlo?.

Calcula el D_3 . ¿Qué significado tiene?

A aquellos pacientes que sienten mejoría todos los días del tratamiento se les realiza un estudio sobre el tiempo de reacción del medicamento (en minutos), encontrándose recogido los datos en la siguiente tabla:

Tiempo de reacción	Nº de pacientes
0-10	300
10-20	500
20-30	400

30-40	500
40-60	300

Se pide:

Escribir las fórmulas de las diferentes medias e indicar cuál de las tres te parece más adecuada para aplicar en este ejercicio (Razonadamente).

A todos los pacientes que tardan en reaccionar más de 35' se le aplica el medicamento complementario PAQUENODUELA para acelerar los efectos de SINDOLORCABEZON. Hallar el número de pacientes a los que se le aplica este segundo medicamento.

Estudiar la representatividad del tiempo medio de reacción. ¿Es representativo? ¿ Por qué?

El Gobierno está pensando en introducir un medicamento con las características de SINDOLORCABEZON. Existen en el mercado junto con este dos productos más PALACABEZA y SINJAQUECAHOY. El tiempo medio de reacción de cada uno de ellos es respectivamente 25 y 30 minutos, con una varianza de 200 y 300 minutos². Explica detalladamente que criterio de selección estadístico podría aplicar el Gobierno. Según el criterio anterior que medicamento sería el que pasaría a engrosar la lista de medicamentos de la Seguridad Social.

Ejercicio 2: (Medidas de centralización y dispersión)

La empresa automovilística COCHESALMENDRON ha realizado un control de potencia sobre los 1000 motores diesel que se han fabricado a lo largo del mes de noviembre del año 1997 obteniendo la siguiente tabla:

Potencia en CV	Frecuencias	x_i n_i
0-50	50	
50-60	200	
60-65	400	
65-70	300	
Más de 70	50	4000 ^(*)

Se pide:

Sin utilizar el dato en negrita que aparece en la tabla anterior, ¿podrías representar gráficamente el histograma de frecuencias? ¿Por qué? (Razona detalladamente)

Calcula la potencia mediana de los motores. Sin el dato en negrita no podrías calcular ni la media (¿Por qué?) ni la moda (¿Por qué?), sin embargo calcular ambos promedios haciendo uso del dato en negrita. e indicando que se ha supuesto para estos cálculos.

En la especificación técnica del motor se indica que tiene una potencia mínima de 55 CV. Hallar el porcentaje de motores con una potencia mayor que está (Nota: Realizarlo por dos métodos: Cuartiles y proporcionalidad).

Estudiar la representatividad de la media aritmética. ¿Sería representativa?

Los motores con menos de 55 CV se apartan de los demás y se estudia el número de piezas defectuosa que han motivado la pérdida global de potencia, obteniéndose la siguiente tabla:

Valores x_i	Frecuencias n_i
1	40
2	30
3	20
4	10

¿La media armónica, la media geométrica y la media aritmética guardan alguna relación de orden? Calcular estás medias y compruébalo.

Representa gráficamente los distribución de frecuencias de la tabla.

Calcula la moda y el recorrido intercuartílico.

¿Qué diferencia existe entre subpoblación y encuesta?

¿Según que criterio nos permite diferenciar las características de una población?

Ejercicio 3: (Medidas de centralización y dispersión)

Se ha realizado una estadística en el centro comercial CONTINENTOL sobre los gastos (en miles de pesetas) que una familia tiene cuando realiza sus compras un día cualquiera de la semana. Este estudio nos aporta la siguiente tabla:

Intervalos	Frecuencias
0-5	1000
5-10	1100
10-20	1600
20-50	1000
50-100	300

Se pide:

¿Cuál es el motivo por el que los datos se presentan en intervalos?

¿ Te parece coherentes los datos de la tabla, o bien tendrías que estudiar su procedencia antes de continuar el estudio?

Halla los ingresos que en ese día tuvo el centro comercial y el gasto medio, modal y mediano de cada familia.

Si a todas las familias que gastan más de 40.000 pesetas, se les obsequia con una bolsa de deporte o una cafetera, ambas valoradas en 2.500 pesetas. Hallar el número de regalos que realiza el centro comercial, así como el porcentaje de clientes que se benefician de ellos. (Nota: utilizar percentiles)

Hallar el primer cuartil. ¿Qué significado tiene?

Estudiar la representatividad del gasto medio. ¿Es representativa? ¿ Por qué?

De las diez personas encargadas de realizar la encuesta se sabe, de años anteriores, que tres de ellos se equivocan al elaborar la encuesta un 10% de las veces; cuatro rellenan ellos mismos el cuestionario y se equivocan el 75% de las veces y el resto son muy detallistas y cumplen perfectamente su trabajo. Se pide: ¿Cuál es la variable? ¿Cuál es la frecuencia? ¿Por qué los datos no se dividen en intervalos? Hallar la media de la variable.

Ejercicios 4: (Medidas de centralización y dispersión)

Se realiza una estadística en dos centros de enseñanza, uno público y otro privado, referente a la nota global del bachillerato de cada uno de los alumnos que van a acudir a los exámenes de selectividad. Las distribuciones de frecuencias son las siguientes:

Centro privado

Nota global de cada alumno.	Frecuencias
5,5	10
6.5	15
7.5	20
8.5	30
9.5	15

Centro público

Nota global de cada alumno.	Frecuencias
[5 , 6]	250
(6 , 7]	150
(7 , 9]	100
(9, 10]	20

Se pide:

- A la vista de la tabla, te sugiere algún comentario de especial importancia. ¿Cuál es el motivo de que los datos se presente en dos tablas de diferente tipo ?
- Estudiar las diferentes medidas de tendencia central (promedios) en las dos distribuciones. En cada distribución ¿cuál te parece más representativo? ¿por qué?
- Hallar el porcentaje de alumnos que en cada centro tiene una nota global superior al 7.
- Hallar los cuartiles primero y tercero de las dos distribuciones.
- Estudiar la representatividad de las medias obtenidas en las distribuciones por separado. ¿En cuál de las dos es más representativa?
- Dos alumnos pertenecientes el primero al centro privado y el segundo, al centro público, solicitan una beca para continuar sus estudios en la universidad mejicana de Acapulco. el primero tiene una nota global de un 8.5 y el otro de un 7. Si sólo se concede una beca ¿quién sería el candidato a obtenerla aplicando el criterio estadístico de la variable tipificada?

Ejercicios 5: (Medidas de centralización y dispersión)

A la finalización del curso "Informática e Internet" se realizó un examen tipo test a los 300 alumnos obteniéndose la siguiente tabla relativa al número de preguntas acertadas:

Nº preguntas acertadas	Nº de alumnos
0-10	10
10-15	20
15-20	60
20-23	100
23-25	70
25-30	30
30-40	10

Se pide:

Representa gráficamente la distribución de frecuencias anterior

Hallar la media

¿Cuál será el número de preguntas tal que la mitad de los alumnos obtengan un número de preguntas acertadas mayor que está?

¿Cuál es el número de preguntas que más se repite? Contesta con todo detalle.

Para la realización de la segunda parte del curso se convocan sesenta plazas. Hallar el número de preguntas mínima que ha debido acertar un alumno para poder realizar la continuación del curso.

Una vez finalizado este segundo curso, se realiza un examen a los alumnos obteniéndose las siguientes notas:

Notas	Nº Alumnos
4	8
5	12
5.5	15

6	14
6.5	6
8	5

Se pide:

¿Por qué no se agrupan los datos en intervalos, como anteriormente?

Halla la mediana, la moda, la media armónica y el recorrido intercuartílico.

Hallar y estudiar la media y su representatividad.

¿Qué resulta más meritorio, obtener 28 preguntas acertadas en el primer examen u obtener un 6.5 en el segundo?

Si se concede un 15% de diplomas entre los alumnos de la segunda parte del curso. ¿A partir de qué nota se concederán?

Ejercicio 6: (Regresión y correlación)

Dada la difícil situación por la que atraviesa actualmente la empresa QUEMALAPATA en la que hemos empezado a trabajar, se propone la reducción de determinados gastos. Para ello se estudia la relación que existe entre dos variables como son: los gastos en publicidad (variable X) y los beneficios (variable Y). De ambas variables disponemos de los siguientes datos:

Año	1993	1994	1995	1996	1997
Gastos en Publicidad (10⁵ ptas)	70	75	80	90	104
Beneficios (10⁶ ptas)	33	45	50	65	67

Se pide:

¿Se puede considerar que ambas variables guardan algún tipo de relación?

¿Cuál sería la variable dependiente y cuál la independiente?

Realizando un gráfico adecuado. ¿Se puede suponer que la relación que las liga es de tipo lineal?

Construye las dos rectas de regresión mínimo cuadrática asociada con las variables.

Si la empresa para el próximo año realizará un esfuerzo para poder invertir 11.500.000 ptas en publicidad. ¿Cuáles resultarían ser sus beneficios? ¿Con qué fiabilidad realizaría usted la predicción?

¿Cuáles resultarían ser sus beneficios si la predicción se efectúa considerando tan solo como variable explicativa el tiempo? ¿Cuál sería la fiabilidad de esta otra predicción? Comente los resultados.

Ejercicio 7: (Regresión y correlación)

Un estudiante de Estadística de la provincia de Sevilla, para poder pagarse sus estudios debe trabajar como camarero en un bar de copas de su localidad CASTILLEJADELCUESTON, al cual suelen acudir todos los jóvenes de la zona. Este año con los conocimientos aprendidos decide por fin estudiar la relación existente entre las galletas saladas y el consumo de bebidas, ya que, es costumbre, dar al cliente este aperitivo cuando se pide una consumición.

Las galletas no pueden tener una concentración de sal superior a 35 gramos por cada 10.000 galletas, y por ello decide ir variando a partir de 10 gramos la concentración de 5 en 5 gramos cada semana e ir anotando el incremento en caja semanalmente. Obteniendo la siguiente tabla:

Gramos de sal por cada 1000 galletas	1	1.5	2	2.5	3
Ingresos en caja (en pesetas)	140300	150000	165000	175000	200000

Se pide:

Establecer un modelo lineal que relaciona las dos variables, estudiando la fiabilidad de dicho modelo

Como consecuencia de los resultados anteriores el propietario del bar decide añadir a las galletas 40'25 gramos de sal, que coincide con toda la sal que tiene , en los almacenes . Realiza una predicción de los ingresos en caja y comente el resultado.

Si el propietario desea unos ingresos de 160.000 pesetas que cantidad de sal aportaría a las galletas. Si aporta 2,75, ¿cuál sería el ingreso en caja?.
Explicar cual de las dos predicciones te merece mayor confianza.

Ejercicio 8: (Regresión simple-múltiple y correlación)

Un holding formado por cinco empresas públicas, decide en vista de una próxima privatización realizar un análisis conjunto de sus actividades.

La valoración del conjunto de todas las empresas es de 20 (en miles de millones de pesetas), donde las empresas A y B representan cada una de ellas el 30% de la valoración total. La empresa C supone un 10% de dicha valoración repartiéndose a partes iguales el resto las otras dos empresas.

Se pide:

Hallar el valor medio de las empresas del holding y estudiar si este es representativo.

Si el beneficio neto que obtuvieron las empresas en el periodo anterior fue el siguiente:

Empresa	Beneficios (en millones de pesetas)
A	22'4
B	23'3
C	10'7
D	15'5
E	13'0

Se pide:

¿Es equitativo el reparto de los beneficios entre las empresas?.

Construye un modelo lineal que explique los beneficios en función del valor de la empresa. ¿Es aceptable el modelo construido?

Aparece una nueva variable: número de empleados (en cientos de trabajadores). Sabemos que la recta de regresión del nº del empleados respecto a la valoración es:

$$\text{nº empleados} = 0.53 + 1.07 \text{ Valoración},$$

con un coeficiente de determinación igual a 0'97. También sabemos que el coeficiente de correlación de la recta de regresión del beneficio respecto al número de empleado es igual a 0'95.

Se pide:

Hallar la matriz de varianza-covarianza de las tres variables y el porcentaje de varianza explicada por el plano de regresión que explica el comportamiento del beneficio en función del número de empleados y la valoración de la empresa.

Ejercicio 9: (Regresión simple-múltiple y correlación)

Un hipermercado ha decidido ampliar el negocio. Decide estudiar de forma exhaustiva el número de cjas registradoras que va a instalar, para evitar grandes colas. Para ello, se obtuvieron los siguientes datos procedentes de otros establecimientos similares acerca del número de cajas registradoras (variable X_2) y del tiempo medio de espera (variable X_1).

Nº de cajas registradoras	10	12	14	12	18	20
Tiempo medio de espera	59	51	42	32	26	22

Bajo el supuesto de que el tiempo de espera medio depende linealmente del número de cajas registradoras se pretende saber:

¿Cómo varía el tiempo medio de espera por cada unidad de caja adicional?

Si se instalaran 17 cajas registradoras, ¿Cuál sería el tiempo medio de espera? ¿Es fiable dicho dato?

Posteriormente, se pensó que podía ser conveniente para explicar el tiempo de espera tener en cuenta además el número de clientes que acuden diariamente a estos establecimientos (variable X_3), por lo que se estudió el tema, obteniéndose un coeficiente de correlación lineal entre ambas variables de 0'9343, siendo el número medio de clientes para los establecimientos anteriores de 850 personas, con un coeficiente de variación del 10'74%. Además, la covarianza entre el número de cajas y el número de clientes que acuden diariamente a estos establecimientos resultó ser de -300. Calcule los parámetros del modelo lineal que explica el tiempo medio de espera en función del número de cajas y del número de clientes que acuden diariamente.

Ejercicio 10: (Regresión simple y correlación)

La empresa GURSO, S.A. ha recibido sugerencias para que invierta sus excesos de tesorería en acciones de un poderoso grupo financiero nacional, dado que su cotización

diaria en la Bolsa de Valores nacionales proporciona una gran liquidez a estas inversiones, y la evolución reciente de sus cotizaciones permiten pronosticar interesantes rentabilidades. El Departamento de Análisis Financiero de GURSO, S.A. recibe el encargo de realizar un estudio sobre la rentabilidad a corto plazo de estas acciones.

El Sr. González, Director del Departamento de Análisis Financiero, ha estudiado la rentabilidad de estas acciones utilizando la información sobre las variaciones de las cotizaciones bursátiles, en un ejercicio simulado de cinco operaciones de compraventa de un mismo número de títulos de este grupo en cortos periodos recientes, siempre con referencia a la Bolsa de Valores de Madrid.

Ha observado que los precios en el primer y último periodo eran iguales y un 10% superiores a los precios de los periodos segundo y cuarto, que a su vez eran un 10% inferiores al precio del tercer periodo que es de 1000 pesetas. ¿Se podría suponer que el precio medio resultante de estas operaciones simuladas suficientemente representativo de este conjunto de cinco precios?

Vistas las rentabilidades obtenidas el Sr. González se ha propuesto explicar el beneficio de cada operación simulada mediante una relación lineal del precio de adquisición de las acciones, ya que ha encontrado que para un precio de adquisición de 950 pesetas por acción, se obtiene un beneficio de 400 pesetas en cada título negociado, y que cuando el precio de las acciones se incrementan en 2 pesetas, el beneficio de la operación de compraventa se reduce en 10 pesetas. Los cálculos internos del Departamento de Análisis financiero de GURSO, S.A. muestran que la varianza de los beneficios simulados es 26 veces la varianza de los precios de adquisición. ¿Cuál es la expresión numérica de la función que están manejando en el Departamento de Análisis Financiero de GURSO, S.A.? ¿Cuál es el beneficio medio de las cinco operaciones de compraventa simuladas en dicho Departamento? ¿Puede indicarnos, razonadamente, si dicha función lineal proporciona un grado de explicación aceptable para el beneficio por el precio de adquisición?

La Dirección de GURSO, S.A. sólo está dispuesta a realizar estas inversiones si el beneficio por acción es de, al menos, 400 pesetas. ¿Qué precio máximo es el que determinará el Sr. González para poder recomendar la inversión?

Ejercicio 11: (Regresión simple-múltiple y correlación)

La dirección de la empresa THALESSA ha llevado a cabo un análisis del coste salarial mensual S (datos en 10^4 pesetas) de sus 115 empleados en relación con un determinado indicador de productividad P, habiendo resultado la siguiente tabla de frecuencias conjunta:

Intervalos de clase	Intervalos de clase de P			
de S	6-8	8-10	10-12	12-14

10-12	28	2	0	0
12-15	3	31	1	0
15-18	2	19	2	0
18-22	0	6	6	0
22-28	0	0	8	1
28-38	0	0	0	6

De estos datos se desea obtener respuesta a las siguientes preguntas:

¿Qué porcentaje de sus empleados tiene un coste salarial por encima del coste salarial medio?

Para el próximo ejercicio económico la dirección tiene pensado plantear al Comité Laboral un incremento lineal de 5.000 pesetas mensuales para los salarios más bajos que suponen el 30% del coste total salarial actual. Para el grupo de los restantes empleados desea proponer un incremento salarial proporcional a sus salarios actuales. ¿En qué porcentaje deberían aumentarse los costes salariales de este segundo grupo de modo que ambos criterios den lugar al mismo resultado para el salario que separa ambos grupos? ¿En qué porcentaje se incrementaría el coste total salarial?

¿En qué medida puede pensarse que el coste salarial explica mediante una relación lineal el indicador de productividad?

Supuesto que se ha calculado esta relación lineal, ¿para qué coste salarial se verifica que un incremento del 1% se traduce en un incremento del 0.5% del indicador de productividad?

La dirección dispone además de los datos sobre absentismo laboral de sus empleados, medido por el número de ausencias en los últimos doce meses, y desea introducir esta variable A en su análisis. Los cálculos conducen a las siguientes covarianzas: $Cov(S, A) = -15$, $Cov(P, A) = -5$. Trabajando con un ajuste lineal que explique el indicador de productividad de sus empleados mediante el coste salarial y la medida de si absentismo laboral, observa que manteniendo constante el coste salarial un incremento de una ausencia laboral se traduce en que el indicador de la productividad decrece en 0'41 unidades. Si el número medio de ausencias en un año es de 5'63 ausencias por empleado, ¿cuál es la dispersión relativa de la variable que expresa el absentismo laboral?

