

INTRODUCCIÓN A LA ESTADÍSTICA – PROBLEMAS – TEMA 2

Los problemas marcados con un asterisco deben resolverse utilizando EXCEL.

- 1.** Calcula la media muestral, la mediana muestral, la moda muestral, el primer y tercer cuartil muestrales, el rango muestral y el rango intercuartílico muestral para los siguientes conjuntos de datos:

- a. El número de quejas recibidas en el departamento de atención al cliente de una compañía eléctrica en las últimas diez semanas:

13 15 7 16 8 4 21 12 3 15

- b. Las predicciones para el IPC (en puntos porcentuales) del próximo año de un país realizadas por nueve servicios de estudios diferentes:

3.2 3.1 4.3 3.7 3.5 3.7 3.4 3 3.6

- 2.** Se han observado las ventas (en cientos de euros) en un día de 6 sucursales de una franquicia situadas en la ciudad A, obteniéndose los resultados siguientes:

5 3 19 21 4 14

Se han observado también las ventas (en cientos de euros) en un día de 7 sucursales de esa franquicia situadas en la ciudad B, obteniéndose los resultados siguientes:

8 10 18 14 4 6 17

- a. Llamando X_1, \dots, X_6 a las observaciones de la muestra de la ciudad A y \bar{X} a la media muestral, calcula las desviaciones a la media $X_1 - \bar{X}, X_2 - \bar{X}, X_3 - \bar{X}, X_4 - \bar{X}, X_5 - \bar{X}$ y $X_6 - \bar{X}$, y deduce la varianza de la muestra de la ciudad A.
- b. Llamando Y_1, \dots, Y_7 a las observaciones de la muestra de la ciudad B, calcula: $\sum_{i=1}^7 Y_i$, (es decir, la suma $Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7$), calcula $\sum_{i=1}^7 Y_i^2$ (es decir, la suma $Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2 + Y_5^2 + Y_6^2 + Y_7^2$), y deduce la varianza de la muestra de la ciudad B.

- c. Compara la variabilidad de las dos muestras utilizando como medida de variabilidad la desviación típica muestral.
 - d. Compara la variabilidad de las dos muestra utilizando como medida de variabilidad el rango intercuartílico muestral.
 - e. Explica en qué situaciones podría ocurrir que las medidas de variabilidad utilizadas en los apartados **c** y **d** llevaran a conclusiones diferentes.

3. *El gestor de una planta de embotellado de agua mineral, que denominaremos “Planta 1”, quiere asegurarse de que el proceso de embotellado de agua en botellas de 1 litro se realiza de modo satisfactorio. Para ello, ha examinado el volumen de agua (en litros) que realmente contienen 40 botellas envasadas en esta plante. En la hoja “Problema 3” del archivo EXCEL adjunto se encuentran los datos en la columna llamada “Planta 1”.
 - a. Calcula la media, la mediana, la moda, el primer cuartil, el tercer cuartil, el rango, el rango intercuartílico, la varianza, la desviación típica y el coeficiente de variación muestrales, indicando la unidad de medida de cada uno de ellos.
 - b. En otra planta de embotellado, que denominaremos “Planta 6”, se ha realizado un análisis similar, tomando una muestra de 35 botellas. En la hoja “Problema 3” del archivo EXCEL adjunto se encuentran los datos en la columna llamada “Planta 6”.
 - b1. Analiza si la media, la mediana y la moda muestrales obtenidas con los datos de la Planta 6 son similares a la media, la mediana y la moda muestrales obtenidas con los datos de la Planta 1.
 - b2. Compara la variabilidad de las dos muestras utilizando como medida la variabilidad el rango muestral.
 - b3. Compara la variabilidad de las dos muestras utilizando una medida de variabilidad que tenga la misma unidad de medida que las observaciones, y que esté basada en los cuartiles muestrales.
 - b4. Compara la variabilidad de las dos muestras utilizando una medida de variabilidad que tenga la misma unidad de medida que las observaciones, y que esté basada en las desviaciones cuadráticas a la media muestral.

4. Las medidas de variabilidad se utilizan también en ocasiones para comparar desigualdad. En este problema se indica cómo se utilizan estas medidas con este fin.

- a. Para analizar los salarios mensuales de las familias de dos poblaciones diferentes A y B, disponemos de una muestra con 3 observaciones de la población A y de otra muestra con 3 observaciones de la población B. Las observaciones de la muestra de la población A (en euros) son: 500, 1.500, 2.500. Las observaciones de la muestra de la población B (en euros) son: 10.000, 12.000, 14.000.
- a1. Utilizando la desviación típica como medida de variabilidad, ¿en cuál de las dos muestra hay más variabilidad?
- a2. Examinando las dos muestras, ¿piensas que hay mayor desigualdad salarial en la muestra de la población A o en la muestra de la población B? **(No es necesario hacer ningún cálculo para responder a este subapartado; simplemente debes dar tu opinión a la vista de los datos).**
- a3. Una de las medidas que se utilizan habitualmente para comparar desigualdad es el coeficiente de variación: cuanto mayor es el coeficiente de variación se considera que hay más desigualdad. Utilizando este criterio, indica cuál de las dos muestras hay más desigualdad salarial.
- b. La tabla siguiente muestra la media y la desviación típica de los salarios mensuales (en euros) de familias de diferentes países.

País	Media	Desviación Típica
A	750	250
B	2000	500
C	1250	750
D	2500	1250

Utilizando el coeficiente de variación como medida de desigualdad, ¿en cuál de estos países hay un reparto más igualitario de los salarios?; ¿en cuál hay un reparto menos igualitario?

5. La profesora Martínez imparte clase en cuatro grupos diferentes. Los estudiantes de cada uno de estos grupos han evaluado (con una nota entre 0 y 10) la actividad docente de esta profesora. Los estudiantes del grupo A le han dado una nota de 6.8, los del grupo B una nota de 6.2, los del grupo C una nota de 5.6 y los del grupo D una nota de 5.8.

- a. La unidad de Calidad Docente del centro otorga a un profesor un “certificado de actividad docente satisfactoria” si la media de las notas recibidas en los grupos que imparte es igual o superior a 6. ¿Conseguirá la profesora Martínez este certificado?
- b. El responsable de la Unidad de Calidad Docente está considerando la posibilidad de modificar el criterio utilizado para otorgar este certificado, y darlo a un profesor cuando la media ponderada de las notas recibidas en los grupos que imparte sea igual o superior a 6, obteniendo esta media ponderada con coeficientes de ponderación proporcionales al número de alumnos de cada grupo.

b1. El número de alumnos en los grupos impartidos por la profesora Martínez ha sido el siguiente: 20 alumnos en el grupo A, 50 en el grupo B, 100 en el grupo C y 30 en el grupo D. Utilizando el nuevo criterio que está considerando el responsable de la unidad, ¿conseguiría la profesora Martínez el certificado de actividad docente satisfactoria?

b2. El profesor López ha impartido clase en los mismos grupos que la profesora Martínez. Los estudiantes del grupo A le han dado al profesor una nota de 7, los del grupo B una nota de 5.5 y los del grupo C una nota de 5. No se conoce todavía la nota dada por los alumnos del grupo D. ¿Cuál es la mínima nota que deben dar los alumnos del grupo D al profesor López para que éste consiga el certificado de actividad docente satisfactoria, se este se otorga utilizando el nuevo criterio que está considerando el responsable de la unidad?

6. Un editor está analizando el número de erratas que contienen los libros publicados por su editorial, por lo que ha analizado las erratas contenidas en una muestra de 400 libros publicados por la editorial. En 119 de estos libros no se detectó ninguna errata, en 158 se detectó una errata y en los 123 restantes se detectaron dos erratas.

- a. Calcula la media muestral del número de erratas.
- b. Calcula la desviación típica muestral del número de erratas.
- c. Calcula el rango intercuartílico muestral del número de erratas.
- d. En los apartados anteriores, ¿se han calculado la verdadera media muestral, la verdadera desviación típica muestral y el verdadero rango intercuartílico muestral, o solo una aproximación de estas medidas muestrales? Justifica tu respuesta.

7. *En la hoja “Problema 7” del archivo EXCEL adjunto se indica el número de accidentes de tráfico que han tenido diferentes clientes de una compañía de seguros en el último año en varias poblaciones.

- a. Para la muestra de la Población 4, calcula la media, la mediana, la desviación típica y el rango intercuartílico muestrales, e indica la unidad de medida de cada uno de ellos.
- b. En una población diferente, que llamaremos Población 7, se ha seleccionado una muestra de 6.000 clientes; de estos 4.000 no tuvieron ningún accidente en el último año, 1.500 tuvieron un accidente, 300 tuvieron dos accidentes y 200 tuvieron tres accidentes. **(En el archivo EXCEL adjunto no hay ningún dato de la Población 7; la única información sobre la muestra de la Población 7 de que se dispone es la que se indica en este enunciado).**

b1. Analiza si la muestra de la Población 7 tiene más variabilidad que la muestra de la Población 4, utilizando una medida de variabilidad basada en los cuartiles muestrales.

b2. Analiza si la muestra de la Población 7 tiene más variabilidad relativa que la muestra de la Población 4, utilizando una medida de variabilidad que esté basada en las desviaciones cuadráticas a la media muestral y que indique variabilidad en proporción a la media.

8. En el año 1990 se obtuvo una muestra de 200 hombres que habían sido padres por primera vez ese año, y se les preguntó su edad. La tabla siguiente muestra los resultados obtenidos, agrupados en intervalos:

Edad	Frecuencia
(15, 25]	110
(25, 35]	80
(35, 45]	10

En el año 2010 se obtuvo una muestra similar, obteniéndose los resultados siguientes:

Edad	Frecuencia
(15, 25]	25
(25, 35]	155
(35, 45]	15
(45, 55]	5

Con estos datos un analista afirma que ha aumentado la variabilidad de la edad a la que los hombres tienen su primer hijo, puesto que la desviación típica aproximada que se obtiene con los datos de 2010 es superior a la desviación típica aproximada que se obtiene con los datos de 1990. ¿Es correcta la afirmación de este analista? Justifica tu respuesta.

9. * En la hoja “Problema 9” del archivo EXCEL adjunto se indican los salarios mensuales (en euros) de muestra de varias poblaciones, en todos los casos agrupadas en 5 clases de igual longitud. Todas las preguntas que van a continuación se refieren exclusivamente a la muestra de la Población 5.

- a.** Representa el histograma de los salarios mensuales de esta muestra.
- b.** Indica razonadamente qué información aporta el histograma realizado en el apartado anterior.
- c.** Indica cuáles son las frecuencias acumuladas relativas de cada una de las clases, y representa el polígono de frecuencias acumuladas relativas.
- d.** Calcula aproximadamente la mediana muestral y el rango intercuartílico muestral.
- e.** Calcula aproximadamente la media muestral, la varianza muestral, la desviación típica muestral y el coeficiente de variación muestral.
- f.** Otra investigadora ha realizado este mismo análisis con los mismos datos, pero midiendo los salarios mensuales en dólares, y utilizando como tipo de cambio 1 euro = 1.09 dólares. Indica cuáles han sido los valores aproximados de la mediana muestral, el rango intercuartílico muestral, la media muestral, la varianza muestral, la desviación típica muestral y el coeficiente de variación muestral que ha obtenido esta investigadora.

10.* Una empresa dedicada a la venta de electrodomésticos quiere analizar la relación entre el gasto en publicidad y las ventas totales. Para ello, dispone de observaciones sobre los gastos en publicidad (en miles de euros) en el último periodo en sucursales pertenecientes a diferentes países, así como las ventas (en miles de euros) en esas mismas sucursales en ese periodo. En la hoja “Problema 10” del archivo EXCEL adjunto se encuentran las observaciones correspondientes. Todas las preguntas que van a continuación se refieren exclusivamente a la muestra de sucursales del País 2.

- a.** Elabora un gráfico que permita analizar la relación entre ambas variables e indica qué información nos aporta el gráfico realizado; en el gráfico, cada eje debe incluir un título con el nombre de la variable representada en él.

- b.** Calcula la covarianza muestral e indica qué información nos aporte el valor obtenido.
- c.** Calcula el coeficiente de correlación muestral e indica qué información nos aporta el valor obtenido.
- d.** Otro investigador ha realizado este mismo análisis con los mismos datos, pero midiendo los gastos en publicidad y las ventas en euros. Indica cuáles han sido los valores de la covarianza muestral y del coeficiente de correlación muestral que ha obtenido este investigador.
- e.** Se sabe que la sucursal catalogada con el número 22 está situada en una zona poco comercial, lo que no ocurre con el resto de las sucursales. Repite los apartados **b** y **c** sin utilizar los datos correspondientes a la sucursal 22, y comenta los resultados.

11. Se ha seleccionado una muestra de 7 estudiantes de la Facultad de Derecho, y se les ha preguntado el tiempo (en horas) X que dedicaron al estudio la semana pasada y el tiempo (en horas) Y que dedicaron a actividades deportivas. Con el fin de estudiar la relación entre X e Y a partir de esta muestra, un investigador ha elaborado la tabla siguiente:

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})(Y_i - \bar{Y})$
18	7	a	b	c
16	5	0	-4	0
19	13	3	4	12
12	9	-4	0	0
26	10	10	1	10
13	8	-3	-1	3
8	11	-8	2	-16

- a.** Determina cuáles son los valores a , b , c que faltan en esta tabla.
- b.** Analiza si X tiene más variabilidad que Y utilizando como medida de variabilidad la varianza muestral.
- c.** Calcula la covarianza muestral entre X e Y .
- d.** Basándote en esta muestra, ¿crees que hay una relación lineal fuerte entre X e Y ? Justifica tu respuesta.

12. Una librería ofrece en internet un catálogo de libros a precio reducido. El propietario de la librería quiere analizar si hay relación entre el precio de un libro y el tiempo que pasa hasta que el libro es vendido. Para hacer este análisis dispone de una muestra con los precios de 151 libros y el tiempo que cada uno de ellos estuvo disponible en internet antes de su venta.

- a. Si llamamos X_1, \dots, X_{151} a los precios en euros de los 151 libros, e Y_1, \dots, Y_{151} a las horas que estos libros estuvieron disponibles en internet antes de su venta, sabemos que $\sum_{i=1}^{151} X_i = 906$, $\sum_{i=1}^{151} Y_i = 7550$ y $\sum_{i=1}^{151} X_i Y_i = 39300$. Con esta información, calcula la covarianza muestral entre estas dos variables.
- b. Indica qué información aporta el valor calculado en el apartado anterior.
- c. Sabemos también que $\sum_{i=1}^{151} X_i^2 = 6786$ y que $\sum_{i=1}^{151} Y_i^2 = 450100$. Con esta información, calcula el coeficiente de correlación muestral entre estas dos variables.
- d. Indica qué información aporta el valor calculado en el apartado anterior, y explica a qué podría deberse la conclusión obtenida.
- e. Si hubiéramos calculado la covarianza muestral y el coeficiente de correlación muestral con estas mismas observaciones, pero considerando el precio de los libros en céntimos de euro y el tiempo en minutos, ¿qué valores habríamos obtenido?

accadem
Universidad