

(P1) ESTADÍSTICA DESCRIPTIVA CON JAMOVİ

INTRODUCCIÓN

[01] Acceda a la página web <https://www.jamovi.org>.

[02] Haga clic en el botón **Jamovi Desktop** para acceder a la descarga.

[03] Descargue el ejecutable **solid**.

[04] Siga todos los pasos del asistente de instalación pulsando **Siguiente**.

[05] Una vez finalizada la instalación, abra Jamovi haciendo clic en su icono ubicado en el sitio que se haya especificado. Una vez abierto, se muestra una ventana con características similares a una hoja de cálculo de Microsoft Excel.

[06] Acceda al **Menú de configuración** pulsando sobre su icono con forma de 3 puntos verticales ubicado en la esquina superior derecha de la ventana de Jamovi.

[07] En la zona **Results** active la opción **3dp** en los campos **Number format** y **p-value format**. [Esta opción restringe el número de decimales permitidos en los resultados del análisis estadístico. ES MUY IMPORTANTE REALIZAR ESTE PASO PARA QUE LOS RESULTADOS DEL ANÁLISIS ESTADÍSTICO TENGAN ESTE GRADO DE PRECISIÓN.]

CASO 1

Los siguientes datos representan las puntuaciones (medidas de 50 a 100), obtenidas por 12 alumnos en las asignaturas de Estadística y Matemáticas que se imparten en el primer curso de Ingeniería:

	Sexo	Estadística	Matemáticas		Sexo	Estadística	Matemáticas
1	Hombre	65	85	7	Mujer	65	94
2	Hombre	50	74	8	Hombre	70	93
3	Mujer	55	76	9	Mujer	55	81
4	Mujer	65	90	10	Mujer	70	91
5	Hombre	55	85	11	Hombre	50	76
6	Mujer	70	87	12	Mujer	55	74

DATOS AUXILIARES:

$$\sum_{i=1}^{12} E_i = 725, \quad \sum_{i=1}^{12} E_i^2 = 44475, \quad \sum_{i=1}^{12} M_i = 1006, \quad \sum_{i=1}^{12} M_i^2 = 84950, \quad \sum_{i=1}^{12} E_i M_i = 61335.$$

[08] Descargue de la plataforma Moodle el fichero **DataBase1**.

[09] En Jamovi, haga clic en la opción **Special Import** del menú principal que está situado en la esquina superior izquierda con forma de 3 líneas horizontales.

[10] En la ventana que se muestra pulse el botón **Browse**.

[11] Localice el fichero descargado en el directorio donde se ubique y pulse **Abrir**. El fichero será importado automáticamente en Jamovi. [Es importante que el fichero origen tenga los datos bien distribuidos, con las variables escritas en columnas. De lo contrario, Jamovi puede tener problemas a la hora de importarlo o bien el fichero de datos resultante puede contener errores. Por ello, NO MODIFIQUE EL FICHERO **DataBase1**.]

[12] Observe el fichero ya importado en Jamovi y tenga en cuenta la información que contiene:

- **ID** (*ID, Text*): número de estudiante.
- **SEXO** (*Nominal, Text*): sexo del estudiante. Los valores numéricos que aparecen corresponden al siguiente código: 1 → HOMBRE, 2 → MUJER.
- **ESTADISTICA** (*Continuous, Integer*): calificación en Estadística.
- **MATEMATICAS** (*Continuous, Integer*): calificación en Matemáticas.

[13] Al importar el fichero Jamovi, este puede conservar las variables A, B y C creadas por defecto. Si

es así, seleccione la cabecera de las columnas A, B y C y pulse en la pestaña **Data** el botón **Delete** de la zona **Variables**. La(s) variable(s) seleccionada(s) será(n) eliminada(s). [Al abrir Jamovi, aparecen por defecto las variables A, B y C para introducir datos manualmente. Es conveniente eliminarlas si, como es nuestro caso, no se van a utilizar. ES MUY IMPORTANTE ELIMINAR VARIABLES DUPLICADAS O ERRÓNEAS DEL FICHERO PARA NO PERJUDICAR EL ANÁLISIS ESTADÍSTICO POSTERIOR.]

[14] Haga doble clic sobre la cabecera de la columna **ID**. En la parte superior aparecerá una ventana con todas sus propiedades. Compruebe que satisface el tipo de variable y el tipo de datos especificado en el paso [12]. [Siempre que una variable contiene simples datos (es decir, no calculados con una fórmula o transformados de otra variable), Jamovi le otorga dos características fundamentales:

- **Measure type**: indica el tipo de variable. Los cuatro tipos posibles son **Nominal** (variable categórica sin jerarquía entre las categorías), **Ordinal** (variable categórica con jerarquía entre las categorías), **Continuous** (variable numérica) e **ID** (variable que contiene etiquetas que designan cada dato individual. No forman parte del análisis estadístico).

- **Data type**: indica el tipo de dato que está recogiendo la variable. Los tres tipos posibles son **Integer** (números enteros), **Decimal** (números decimales) y **Text** (texto).]

[15] Repita el paso [14] con el resto de variables del fichero.

[16] La variable **SEXO** es numérica al principio. Al cambiar sus características, aparecerá la lista **Levels**, rellena con los valores numéricos que tiene por defecto. Cambie los valores por los códigos indicados en [12]. Estos cambios se reflejarán automáticamente en todo el fichero. [NO CAMBIE LOS VALORES DE FORMA MANUAL TECLEANDO SOBRE EL FICHERO DIRECTAMENTE, ya que si comete un error el análisis estadístico puede verse afectado.]

PREGUNTA 1.- ¿Qué nota es la más homogénea?

Respuesta manual:

1

[17] En la pestaña **Analyses**, haga clic en la opción **Descriptives** del módulo **Exploration**.

[18] Seleccione las variables **ESTADISTICA** y **MATEMATICAS** de la lista y pulse el botón '→' para añadirlas a la lista **Variables**. [Para quitar una variable de la lista **Variables** que no se quiera analizar, selecciónela y pulse el botón '←'.]

[19] Despliegue el menú **Statistics** y seleccione las medidas descriptivas que quiera calcular sobre las variables. Todas las seleccionadas se mostrarán en la ventana **Results**. [Las medidas descriptivas incluidas en el menú **Statistics** son, entre otras:

- **N**: total de datos.

- **Missing**: total de valores ausentes.

- **Mean**: media aritmética $\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j$.

- **Median**: mediana, valor x_j que deja a cada lado el 50% de los datos.

- **Standard deviation**: cuasi-desviación típica $\bar{s} = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2}$

- **Minimum**: dato x_j de valor mínimo.

- **Maximum**: dato x_j de valor máximo.

- **Mode**: dato x_j que más se repite.

- **Sum**: suma de todos los datos $\sum_{j=1}^N x_j$.

- **Percentiles**: un percentil p_k (siendo k un entero entre 1 y 99) es un dato x_j que deja a su izquierda el $k\%$ de los datos. Escriba en este campo uno o más enteros k para calcular los respectivos percentiles. Por defecto aparecen escritos los números 25, 50 y 75, que corresponden a los percentiles 25 (primer cuartil), 50 (mediana) y 75 (tercer cuartil).

- **Variance**: cuasi-varianza $\bar{s}^2 = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})^2$.

- **Range**: rango de datos $R = x_{\max} - x_{\min}$.

- **IQR**: rango intercuartílico $RQ = p_{75} - p_{25}$.

◦ **Skewness**: coeficiente de asimetría $A = \frac{1}{N s^3} \sum_{j=1}^N (x_j - \bar{x})^3$. Mide la desviación horizontal de los datos con respecto a una distribución normal. Si $A > 0$, la distribución de los datos es asimétrica a la derecha; si $A < 0$, la distribución es asimétrica a la izquierda.

◦ **Kurtosis**: coeficiente de curtosis $K = \frac{1}{N s^4} \sum_{j=1}^N (x_j - \bar{x})^4 - 3$. Mide la desviación vertical de los datos con respecto a una distribución normal. Si $K > 0$, la distribución de datos es más apuntada que una normal o leptocúrtica; si $K < 0$, la distribución es más plana que una normal o platocúrtica.

◦ **Std. error of Mean**: error estándar de la media $s_{\bar{x}} = \bar{s}/\sqrt{N}$. Mide la desviación típica de todas las medias que pueden tomarse con todas las muestras de tamaño N posibles de la población de la cual se han obtenido los datos.]

Compruebe que el resultado de la **PREGUNTA 1** en Jamovi coincide con el calculado manualmente.

PREGUNTA 2.- ¿Están relacionadas ambas notas?

Respuesta manual:

[20] En la pestaña **Analyses** haga clic en la opción **Correlation Matrix** del módulo **Regression**.

[21] Seleccione las variables **ESTADISTICA** y **MATEMATICAS** y añádalas a la lista de la derecha con el botón '→'.

[22] Active la casilla **Pearson**. Se mostrará en la ventana **Results** el coeficiente de correlación lineal de Pearson ρ entre las variables. [Para medir el grado de relación lineal entre dos variables **Continuous**, se recurre al coeficiente de correlación lineal de Pearson ρ . Cuanto más cerca esté este coeficiente de 1 (resp. -1), mayor relación lineal directa (resp. inversa) habrá entre las variables; cuanto más cerca esté de 0, mayor será su independencia o su relación de otro tipo que no sea lineal (hiperbólica, logarítmica, potencial, exponencial, etc.)]

Compruebe que el resultado de la **PREGUNTA 2** en Jamovi coincide con el calculado manualmente.

PREGUNTA 3.- Determine la puntuación en Matemáticas que se espera para un alumno que en Estadística ha obtenido 62 puntos. ¿Qué porcentaje de la variación de la nota de Matemáticas se explica con el modelo calculado?

Respuesta manual:

[23] En la pestaña **Analyses**, haga clic en la opción **Linear Regression** del módulo **Regression**.

[24] Seleccione la variable **ESTADISTICA** y añádala al campo **Dependent Variable** con el botón '→'.

[25] Seleccione la variable **MATEMATICAS** y añádala al cuadro **Covariates** con el botón '→'. Se mostrará en la ventana **Results** una primera tabla que describe el modelo de regresión y a continuación una segunda tabla que contendrá información sobre la bondad de ajuste del modelo. [En general, un modelo de regresión lineal expresa una variable independiente Y en función de una serie de variables independientes X_1, X_2, \dots, X_k a través de una fórmula del tipo $\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$. Obtener el modelo supone estimar los coeficientes a, b_1, b_2, \dots, b_k y una vez conseguido, puede utilizarse para realizar predicciones o estimar valores de Y sustituyendo los datos conocidos en las variables X_i . Cuando Jamovi calcula un modelo de regresión, en la primera tabla aparecen los coeficientes estimados b_i del modelo asociados a cada variable y por otro lado, el término independiente a (la ordenada en el origen) asociado al nombre **Intercept**. En la segunda tabla aparece la bondad de ajuste del modelo, medida a través del coeficiente de determinación R^2 . Cuanto más cerca esté este coeficiente de 1, mejor será el ajuste del modelo; si está cerca de 0, el modelo no resulta útil.]

Compruebe que el resultado de la **PREGUNTA 3** en Jamovi coincide con el calculado manualmente.

PREGUNTA 4.- Agrupando las puntuaciones de ambas asignaturas en dos intervalos de igual amplitud, obtener las principales medidas estadísticas de posición (media, moda, mediana, primer y tercer cuartiles) y dispersión (cuasi-varianza, cuasi-desviación típica, rango, rango intercuartílico, coeficiente de variación y covarianza) para ambas series de datos, interpretando los resultados.

Respuesta manual: 4

[26] Seleccione una columna en blanco al final del fichero de datos y en la pestaña **Data** haga clic en el botón **Compute**. Seguidamente haga doble clic sobre la cabecera de dicha columna y en la parte superior aparecerá una ventana para realizar el proceso.

[27] Escriba el nombre **EST AGRUPADA** en el campo **Computed variable**.

[28] Teclee la expresión del cálculo necesaria para la agrupación de los datos en la barra de fórmulas situada justo debajo. Para ello, al hacer clic sobre el botón f_x , se mostrará un desplegable con diferentes opciones de funciones típicas de hojas de cálculo y los nombres de las variables del fichero que se pueden utilizar. En este caso, la expresión sería:

$$\text{IF}(\text{ESTADISTICA} \leq 60, 55, 65)$$

[Para seleccionar las variables del fichero e incorporarlas a la fórmula, haga doble clic sobre su nombre en el desplegable f_x .]

[29] Una vez escrita la expresión, pulse la tecla **Enter** y se creará automáticamente la nueva variable calculada en el fichero con los datos agrupados. [Si denotamos a las clases $C_j = (a_{j-1}, a_j]$, entonces los valores numéricos que aparecen corresponden a las marcas de clase $x_j = (a_{j-1} + a_j)/2$.]

[30] Realice lo mismo con la variable **MATEMATICAS**, calculando una variable llamada **MAT AGRUPADA** a través de la fórmula $\text{IF}(\text{MATEMATICAS} \leq 84, 79, 89)$.

[31] Una vez creadas las variables **EST AGRUPADA** y **MAT AGRUPADA**, repita los pasos [17], [18] y [19] para calcular las medidas estadísticas requeridas. En el caso de la covarianza, puede ser obtenida a partir del cálculo del coeficiente de correlación lineal de Pearson como en los pasos [20], [21] y [22].

Compruebe que el resultado de la **PREGUNTA 4** en Jamovi coincide con el calculado manualmente.

PREGUNTA 5.- ¿Dependen las puntuaciones de Estadística (según la agrupación anterior) del sexo del alumno? En caso afirmativo, ¿en qué grado?

Respuesta manual: 5

[32] En la pestaña **Analyses** haga clic en la opción **Independent Samples** del módulo **Frequencies**.

[33] Seleccione la variable **SEXO** y añádala al campo **Rows** con el botón ' \rightarrow '.

[34] Seleccione la variable **EST AGRUPADA** y añádala al campo **Columns** con el botón ' \rightarrow '.

[35] Despliegue el menú **Statistics** y active las casillas **Contingency coefficient** y **Phi and Cramer's V**. Se mostrará en la ventana **Results** una tabla de contingencia y los valores de los coeficientes χ^2 , Φ , V de Cramer y C de contingencia. [Una tabla de contingencia es una tabla de doble entrada. Sus cabeceras de filas contienen los valores de una variable y sus cabeceras de columnas contienen los valores de otra, por lo que sus elementos centrales lo conforman las frecuencias absolutas de cada pareja posible de valores (uno de cada variable). Las tablas de contingencia únicamente pueden ser calculadas en Jamovi para variables categóricas. Una vez construida, para detectar el grado de asociación entre las variables, siempre que haya al menos una de tipo **Nominal**, se recurre al coeficiente χ^2 de Pearson como

primera medida, y se controla su tamaño normalizándolo, dando lugar a los coeficientes Φ (específico para tablas 2×2) y V de Cramer. Cuanto más cerca estén estos coeficientes de 1 (también -1 en el caso de Φ), mayor será la asociación entre las variables; cuanto más cerca de 0, mayor será su independencia). El coeficiente C de contingencia se interpreta de forma similar al coeficiente V de Cramer.]

Compruebe que el resultado de la **PREGUNTA 5** en Jamovi coincide con el calculado manualmente.

CASO 2

Descargue la siguiente base de datos **Ejemplo1R** disponible en la plataforma Moodle.

Los códigos de las variables **SEXO** Y **LOCALIDAD** son los siguientes:

SEXO:

1 \rightarrow HOMBRE, 2 \rightarrow MUJER

LOCALIDAD:

1 \rightarrow A, 2 \rightarrow B, 3 \rightarrow C.

Para realizar este caso, se empleará el ANÁLISIS POR GRUPOS, es decir, el análisis dividiendo los datos de una variable *Continuous* según las categorías o niveles de una variable *Nominal* u *Ordinal*. Para ello, en la opción *Descriptives* del módulo *Exploration*, además de seleccionar la(s) variable(s) de tipo *Continuous* que quiera, se deberá seleccionar la variable de tipo *Nominal* u *Ordinal* necesaria en la separación por grupos y añadirla a la lista *Split by* con el botón ' \rightarrow '.

Teniendo esto en cuenta, importe el fichero en Jamovi, cambie los valores de los valores **SEXO** y **LOCALIDAD** y responda las siguientes preguntas:

P1.- ¿Qué grupo tiene una talla más dispersa, la de los hombres o la de las mujeres?

6 , ya que el coeficiente de variación de las mujeres es 7 y el de los hombres es 8 .

SEXO	EDAD	TALLA	PESO	LOCALIDAD
1	63	165	70	1
2	55	162	87	1
1	72	165	75	1
2	52	167	76	2
1	54	167	70	2
1	73	170	76	2
1	49	171	95	2
2	51	145	65	2
1	50	178	82	3
2	41	158	75	3
1	42	177	80	1
1	54	170	82	1
2	58	155	62	2
2	38	166	60	2
2	67	152	52	3
2	62	162	87	3
2	45	155	65	3
1	58	173	90	1
2	59	155	60	1
2	48	160	75	1
2	57	152	70	2
1	50	175	83	3
1	58	190	80	3
1	47	173	86	2
2	46	165	82	1
1	40	169	79	1
2	58	157	70	2
2	52	149	56	2
1	45	173	77	3
1	53	167	72	3

P2.- La localidad de moda entre las mujeres es 9 .

P3.- ¿La variable EDAD es simétrica? 10 Simétrica derecha, ya que su coeficiente ϕ_1 tiene un valor de 11 0,443.

P4.- ¿Están relacionadas las variables LOCALIDAD y SEXO? 12 No. ¿Esta relación puede asumirse nula, baja, media o alta? 13 Medio, ya que el coeficiente χ^2 tiene un valor de 14 1,3293, siendo su valor máximo 15 30.

P5.- ¿El modelo de regresión lineal simple para estimar el peso de una persona en función de su edad puede considerarse muy bueno, bueno, malo o muy malo? 16 Muy malo, ya que la bondad de ajuste del mismo es 17 0,01611.

P6.- El 75 % de las personas tiene una talla por debajo de 18 170,8 centímetros.

P7.- La correlación entre la edad de una persona y su talla es 19 -0,122 (negativa).

P8.- ¿La variable PESO puede considerarse que tiene una varianza normal? 20 No, ya que su coeficiente ϕ_2 tiene un valor de 21 -0,33.

P9.- La media de la edad de los hombres que viven en la localidad B es 22 53,9 (split by local sex).

P10.- El rango intercuartílico del peso de las mujeres que viven en la localidad C es 23 6.

P11.- Se pretende obtener el modelo lineal para estimar el peso de una persona en función del resto de variables posibles incluidas en el fichero de datos. En este sentido, la ordenada en el origen de dicho modelo es: 24 cte modelo de regresión -35,874.

P12.- El 50 % de los hombres miden más de 25 53 (mediana = Q_2) kilos.

P13.- Del total de individuos de la localidad C, el porcentaje total de mujeres es 26 44,4.

P14.- Se pretende obtener el modelo lineal simple para estimar el peso de una persona en función de su edad. En este sentido, la pendiente de la recta es 27 (coeficiente que acompaña a x) -0,150.

P15.- ¿Cuál sería el peso estimado, en kilos, de una persona de 22 años de edad? 28 79,936.

P16.- Usando el comando **Expected counts** del menú **Cells** situado en la opción **Independent samples** del módulo **Frequencies**, en el caso de independencia, la frecuencia esperada de mujeres que viven en la localidad A tiene un valor de 29 5.

P17.- La cuasi-varianza del peso es 30 108. ¿Puede decirse sólo con este dato que existe un grado muy alto de dispersión? 31 (con eso sólo no).

P18.- El primer cuartil de la variable TALLA es 32 $Q_1 = 157$.

P19.- El mayor coeficiente de correlación entre las variables del fichero es 33 coeficiente correlación. 0,636

P20.- El porcentaje de individuos que viven en la localidad B es 34 36,7 %.