Introducción al análisis de datos de variables pecuarias

$$\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right] \qquad \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - n\bar{X}^2\right]$$

$$t_c = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{Sp^2 \frac{1}{n1} + \frac{1}{n2}}} \qquad \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Yi^2 - \frac{(\sum_{1=1}^{n} Yi)^2}{n}\right]}$$

$$EED = \sqrt{Sp^2 \frac{n1+n2}{(n1)\,(n2)}} \qquad Sp^2 = \frac{(n_1 - 1){S_1}^2 + (n_2 - 1){S_2}^2}{n_1 + n_2 - 2}$$

Leodan Tadeo Rodríguez-Ortega, Alejandro Rodríguez-Ortega, Filogonio Jesús Hernández-Guzmán, Héctor Leyva Jiménez, Erick Zúñiga Estrada; Sergio Iban Mendoza-Pedroza, Mauricio Velázquez-Martínez, María de la Luz Estrada-Hernández, Jorge San Juan-Lara, Artemio Vargas-Galicia

2023

Universidad Politécnica de Francisco I. Madero

Carretera Tepatepec - San Juan Tepa, kilometro dos, Hidalgo, México

Francisco I. Madero, Hidalgo. México. CP 42660.

Libro:

Introducción al análisis de datos de variables pecuarias

ISBN: 978-607-9260-27-9

Primera edición: junio del 2023

Directorio

Dr. Leoncio Marañón Priego

Rector de la Universidad Politécnica de Francisco I. Madero, Tepatepec, Hidalgo, México

MC. María de la Luz Estrada Hernández

Directora de la Ingeniería en Producción Animal de la UPFIM

L. C. José Erick Juárez Martínez

Jefe de la subdirección de Recursos Materiales

Índice de contenido

Índice de figuras

Índice de cuadros

Presentación

Este libro tiene como finalidad ayudar en el análisis de datos de variables cuantitativas empleadas en la producción animal. Asimismo, un documento de ayuda a los estudiantes que realizan experimentos en ciencias pecuarias.

Leodan…

Introducción

Una de las necesidades de los alumnos que cursan materias relacionadas con la producción animal es poseer sólidos conocimientos en el uso de las herramientas estadísticas. La estadística tiene aplicaciones directas y concretas en la vida real, ya que toma los números y cifras de diferentes fenómenos que suceden en las actividades pecuarias como, por ejemplo: el peso vivo, la ganancia de peso, el consumo de alimento, el porcentaje de nacimientos y muchos otros datos incluso más complejos.

La estadística como ciencia nos permite conocer a un nivel mucho más preciso una sociedad. Se considera un método utilizado para recoger, organizar, concentrar, reducir, presentar, analizar, generalizar y contrastar los resultados numéricos de observaciones directas o indirectas de fenómenos reales, así como de la información obtenida a partir de la experimentación, para estar en condiciones de llevar a cabo tanto evaluaciones como conclusiones adecuadas, y tomar decisiones acertadas y confiables. El objetivo de este trabajo tiene como finalidad introducir al lector al análisis de datos de variables cuantitativas empleadas en la ciencia animal. Asimismo, un documento de ayuda para los estudiantes que realizan experimentos en ciencias pecuarias.

Tema 1. Definiciones iniciales

Variable

Todo aquello que vamos a medir o evaluar en un experimento, es decir una variable es susceptible de medición. Vacca (1999) menciona que una variable son todas las características o propiedades de los objetos de estudio.

¿Qué es medir?

Medir es estimar la magnitud de cierta propiedad de uno o más objetos con ayuda de un sistema métrico específico, con un instrumento de medición, escala de medición y unidades de medición.

Tipos de variables

Cuantitativas: estas variables se pueden medir, y pueden tomar cualquier número, con unidades de medida (mg, cm, m, kg, km, °C, °F, otras; Cienfuegos y Cienfuegos, 2016).

Figura 1. La altura es una medida cuantitativa (Imagen tomada de: https://es.dreamstime.com/el-muchacho-est%C3%A1-midiendo-su-hermana-en-la-altura-de-medida-del-dinosaurio-ejemplo-image151382734).

Cualitativa: son aquellas que no pueden ser medidas, no posee una unidad, sin embargo, pueden representarse únicamente con valores enteros. Un ejemplo se representaría como, rural y urbano, alfabeto y analfabeto, son propiedades expresadas al modo cualitativo (Cienfuegos y Cienfuegos, 2016; Rincón 2017).

Figura 2. El sentido del tacto; con la respuesta frio o caliente, una variable cualitativa (Imagen tomada de: https://www.istockphoto.com/es/vector/fr%C3%ADo-caliente-cubo-de-hielo-fogata-palabras-opuestas-en-ingl%C3%A9s-temperatura-gm1480841995-508359578)

Población

Es el conjunto de personas u objetos de los que se desea conocer algo en particular (en esta se puede realizar una investigación). Una población puede estar constituida por personas, animales, registros médicos o muestras de laboratorio (Gallego, 2004).

Muestra

Cualquier subconjunto de la población. Este subconjunto es muy importante que sea representativo de la población. La muestra es una parte representativa de la población (Rincón, 2017).

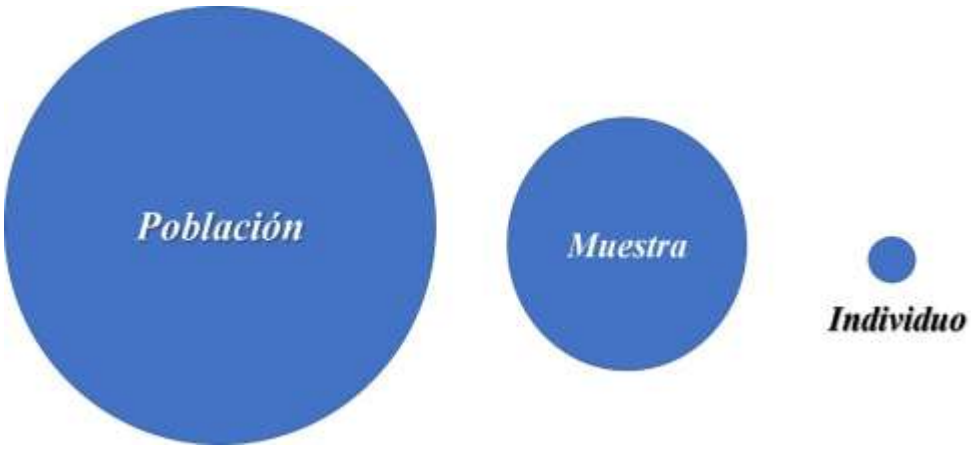


Figura 3. Población, muestra e individuo.

Muestreo aleatorio simple

Se selecciona cierto número de sitios, de cada miembro de cierta población que tenga la misma probabilidad de que sea elegido. Pero esto no garantiza que toda el área de estudio sea cubierta, por las superficies que son relativamente grandes y quedan sin ser muestreadas (Corral *et al.*, 2015).

Muestreo aleatorio estratificado

En este muestreo se realiza la selección de los sitios de muestreo mediante el procedimiento aleatorio simple. En este punto se disminuye la probabilidad de que existan zonas sin muestra o en zonas con una alta concentración de muestras (Corral *et al.*, 2015)

Muestreo sistemático

Esto consiste en elegir los puntos de cada estrato como el sitio de lo que es la muestra. Este brinda un cierto número lo suficientemente observable, que son separadas con las distancias y las direcciones bien definidas (Corral *et al*., 2015).

Muestreo anidado

Se requiere que la población sea dividida por bloques que serán subdivididos en otros más definidos, hasta que se alcance el nivel de detalle que se desee saber. En cada uno de los niveles, los bloques se anidarán en bloques de nivel superior (Corral *et al*., 2015).

Literatura citada

Corral Y, Corral I, Corral AF. 2015. Procedimiento. Revista ciencias de la educación 26 (46): 151-167.

Cienfuegos VM de A, Cienfuegos VA. 2016. Lo cuantitativo y cualitativo en la investigación. Un apoyo a su enseñanza. Revista Iberoamericana para la Investigación y el Desarrollo Educativo 7 (13). URL: https://www.ride.org.mx/index.php/RIDE/article/view/231/1059

Gallego FC. 2004. Cálculo del tamaño de la muestra. Matronas profesión 5 (18): 5-13.

Rincón L. 2017. Estadística descriptiva. Facultad de Ciencias. Universidad Autónoma de México, Ciudad de México. 210 pp.

Vacca DG. 1999. Definiendo las variables. Odontología Sanmarquina 1 (4): 61-62

Tema 2. Variables de tendencia central

Medía de la muestra también llamada promedio o media, de un conjunto infinito de números es el valor característico de una serie de datos cuantitativos.

Donde:

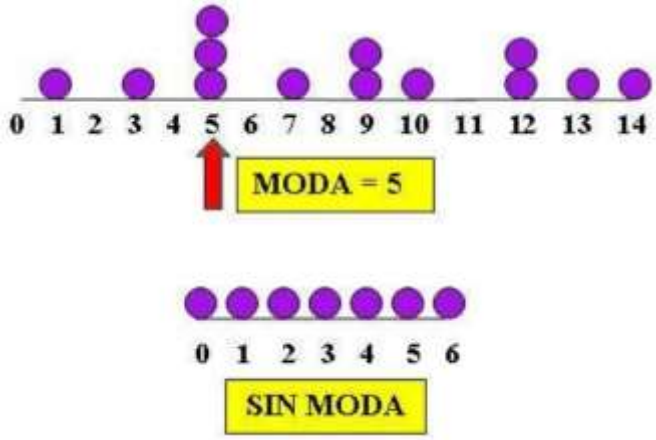
$\overline{X}$ = Media de la muestra (promedio)

$\sum_{i=1}^{n} Xi$ = Sumatoria de Xi

n = número de caso o tamaño de muestra

$$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n}$$

Moda: es el valor con mayor frecuencia en una de las distribuciones de datos.



Mediana: Es el valor que encontramos a la mitad del conjunto de datos.

Mediana (Es el dato central)

4, 5, 6, 6, 6, 7, 7, 8, 8, 9, 9, 10, 11

La mediana en un número impar de los datos: aquí la mediana será la que **ocupará la posición central**. Ejemplo, si los valores de la variable fueran los siguientes.

$$\{9, 10, \mathbf{11}, 12, 13\}$$

Su mediana seria M= 11

La mediana en un número par de los términos: esto ocasionara dos términos centrales y se tomaría como medida, la media aritmética de estos. En este ejemplo, si los valores de la variable son:

$$\{11, 12, 15, \mathbf{17}, \mathbf{19}, 20, 23, 24\}$$

La mediana será = $\frac{17+19}{2} = \mathbf{18}$

Tema 3. Variables de dispersión

Varianza: es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Su símbolo es $\mathbf{S^2}$ si se está trabajando con una muestra y o^2 si se trata de una población (Moncada *et al.*, 2002).

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Dónde:

S^2 = Varianza

X_i = Término del conjunto de datos

$\bar{X}$ = Media de la muestra

$\sum_{i=1}^{n} Xi = Sumatria$

n = Tamaño de la muestra

Para encontrar el valor de la varianza de una muestra se describen tres fórmulas:

Fórmula 1:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]$$

Donde:

$\frac{1}{n-1}$ = **Uno sobre los g*rados de libertad***

$\sum_{i=1}^{n} Xi^2$ = ***Suma de cuadrados no corregidos***

$\frac{(\sum_{1=1}^{n} Xi)^2}{n} = Factor\ de\ corrección$

$\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right] = Suma\ de\ cuadrados\ corregidos$

Fórmula 2:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - n\bar{X}^2\right]$$

Donde:

$\frac{1}{n-1}$ = **Uno sobre los g*rados de libertad***

$\sum_{i=1}^{n} Xi^2 = Suma\ de\ cuadrados\ no\ corregidos$

$n\bar{X}^2 = Factor\ de\ correccion$

$\left[\sum_{i=1}^{n} Xi^2 - n\bar{X}^2\right] = Suma\ de\ cuadrados\ coregidos$

Fórmula 3:

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} (Xi - \bar{X})^2\right]$$

Donde:

$\frac{1}{n-1}$ = **Uno sobre los g*rados de libertad***

$[\sum_{i=1}^{n}(\mathbf{Xi} - \bar{\mathbf{X}})^2] = Suma\ de\ Cuadrados\ corregidos$

Desviación estándar: Es la raíz cuadrada de la varianza (Salazar y del Castillo 2018).

Es importante recordar que la desviación estándar (**S**) tiene las mismas unidades que la variable de interés por ese motivo se explica de mejor manera la dispersión de los datos.

Se expresa en las siguientes fórmulas:

$$s = \sqrt{s^2} = \frac{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}{n-1}$$

La siguiente fórmula se encuentra en el libro: Said Infante Gil y Guillermo P. Zarate de Lara. Métodos Estadísticos un enfoque interdisciplinario. Editorial La Gaya Ciencia Vol.1. 610 pp.

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]}$$

La presente fórmula se encuentra en la página seis del libro bioestadística de Pastor-Barriuso *et al.* (2012).

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - n\bar{X}^2\right]}$$

La siguiente fórmula se encuentra en la página 51 del libro bioestadística de Pastor-Barriuso *et al.* (2012).

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} (Xi - \bar{X})^2\right]}$$

Error estándar o error típico: Es la medida del error que se comete al tomar las mediciones, realizar los cálculos, de una muestra. Es el valor que se cuantifica, cuando se apartan los valores de la media poblacional (Abraira, 2002).

$$E.E. = \frac{S}{\sqrt{n}}$$

Las siguientes fórmulas fueron generadas de las fórmulas citadas en la desviación estándar:

Fórmula 1:

$$\boldsymbol{E.E.} = \frac{\sqrt{\frac{\mathbf{1}}{\boldsymbol{n-1}}\left[\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Xi^2} - \frac{(\sum_{\mathbf{1=1}}^{\boldsymbol{n}} \boldsymbol{Xi})^2}{\boldsymbol{n}}\right]}}{\sqrt{\boldsymbol{n}}}$$

Fórmula 2:

$$\boldsymbol{E.E.} = \frac{\sqrt{\frac{\mathbf{1}}{\boldsymbol{n-1}}\left[\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Xi^2} - \boldsymbol{n\bar{X}^2}\right]}}{\sqrt{\boldsymbol{n}}}$$

Fórmula 3:

$$\boldsymbol{E.E.} = \frac{\sqrt{\frac{\mathbf{1}}{\boldsymbol{n-1}}\left[\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{Xi} - \boldsymbol{\bar{X}})^2\right]}}{\sqrt{\boldsymbol{n}}}$$

Coeficiente de variación

El coeficiente de variación se puede clasificar como una calificación, la cual permite a los usuarios la evaluación de la calidad estadística de las estimaciones. Si esta es muy elevada, significa que está entre las mediciones existe una elevada variación.

$$C.V. = \frac{S}{\bar{X}} * 100$$

Las siguientes fórmulas fueron generadas de las fórmulas citadas en la desviación estándar y la media:

Fórmula 1

$$C.V. = \frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n}(Xi - \bar{X})^2\right]}}{\frac{\sum_{i=1}^{n} Xi}{n}} * 100$$

Fórmula 2

$$\boldsymbol{C.V. = \frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}\right]}}{\frac{\sum_{i=1}^{n} Xi}{n}} * 100}$$

Fórmula 3:

$$C.V. = \frac{\sqrt{\frac{1}{n-1}\left[\sum_{i1}^{n} Xi^2 - n\bar{X}^2\right]}}{\frac{\sum_{i=1}^{n} Xi}{n}} * 100$$

Estudio de caso: "El coeficiente de variación en el peso de bovinos y cerdos"

Un granjero desea conocer ¿cuál de las siguientes especies presenta el mayor coeficiente de variación?, el peso de los bovinos o el peso de los cerdos.

Animales	$\bar{X}$	**S**
Bovinos	750	65
Cerdos	45	35

Operaciones:

Para los bovinos C.V. = $\frac{65}{750} * 100 = 0.08*100=$ Es decir que es el 8%

Para los cerdos C.V. = $\frac{35}{45} * 100 = 0.77*100=$ Es decir que es el 77%

Animales	$\bar{X}$	S	C.V.
Bovinos	750	65	8%
Cerdos	45	35	77%

Conclusión: Con las operaciones anteriores se demostró que el coeficiente de variación en el peso en los cerdos (77%) es mayor al coeficiente de variación en el peso de los bovinos (8%).

Literatura citada

Abraira V. 2002. Desviación y error estándar. SEMERGEN: 28 (11):621-3.

Moncada JJ, Solera HA, Salazar RW. 2002. Fuentes de varianza e índices de varianza explicada en las ciencias del movimiento humano. Revista de Ciencias del Ejercicio y la Salud 2 (2): 70-74.

Pastor-Barriuso R. 2012. Bioestadística. Centro Nacional de Epidemiología Instituto de Salud Carlos III. 251 pp. I.S.B.N.: 978-84-695-3775-6.

Salazar PC, del Castillo GS. 2018. Fundamentos básicos de estadística. Quito, Ecuador, Primera edición. 224pp. ISBN: 978-9942-30-616-6.

Infante GS., Zarate de Lara GP. 2008. Métodos Estadísticos un enfoque interdisciplinario. Editorial La Gaya Ciencia Vol.1. 610 pp.

Tema 4. Procedimientos de SAS para el análisis de datos [variables de tendencia central y de dispersión]

PROC PRINT es un procedimiento que se utiliza para enlistar e imprimir los datos. Este procedimiento es utilizado con el objetivo de revisar que los datos se han leídos correctamente en el programa SAS.

Ejemplo:

Un conjunto de datos de la variable X, se desean imprimir para observar que los datos están escritos de forma correcta.

```
Title"Clase uno";
Data Uno;
Input X;
Cards;
50
40
22
17
99
15
60
7
77
12
97
69
;
Proc Print;
Run;
```

Resultados de la salida

```
                                  Clase uno
                                  Obs     X
                                    1    50
                                    2    40
                                    3    22
                                    4    17
                                    5    99
                                    6    15
                                    7    60
                                    8     7
                                    9    77
                                   10    12
                                   11    97
                                   12    69
```

PROC MEANS tiene como objetivo analizar las variables de tendencia central y de dispersión de un conjunto de datos. La sentencia **Var** se incluye en la variable de respuesta cuya medida se quiere obtener X.

Significado de las abreviaciones en el procedimiento:

- Proc Means: Procedimiento
- Mean: Media
- STD: varianza
- STDERR: Desviación estándar
- CV: Coeficiente de variación
- Min: Valor mínimo
- Max: Valor máximo

Ejemplo: Se desea conocer las variables de tendencia central y de dispersión del siguiente conjunto de datos (X).

```
Title"Clase uno";

Data dos;
Input X;
Cards;
50
40
22
17
99
15
60
7
77
12
97
69
;
Proc Means Mean STD STDERR CV MIN MAX;
Var X;
Run;
```

Resultados de la salida

Clase uno
Procedimiento MEANS

Variable de análisis: X

Media	Desviación estándar	Error estándar	Coeficiente de variación	Mínimo	Máximo
47.0833333	33.2878857	9.6093849	70.6999342	7.00	99.00

Conclusión: El procedimiento Proc Means es de bastante ayuda en el análisis de datos, sin embargo, es necesario conocer y escribir los comandos que darán la orden del análisis

PROC UNIVARIATE ayuda en el análisis más profundo de los datos; explorar las variables y obtener estadísticas descriptivas, así como la distribución de los datos. Ejemplo: Se desea conocer las variables de tendencia central y de dispersión del siguiente conjunto de datos (X). El siguiente ejemplo se puede copiar directamente al programa SAS.

```
Title"Clase uno";
Data tres;
Input X;
Cards;
50
40
22
17
99
15
60
7
77
12
97
69
;
Proc Univariate;
Var X;
Run;
```

Resultados de la salida

Procedimiento UNIVARIATE

Momentos

N	12	Pesos de la suma	12
Media	47.0833333	Observaciones de la suma	565
Desviación típica	33.2878857	**Varianza**	1108.08333
Suma de cuadrados no corregidos	38791	Suma de cuadrados corregidos	12188.9167
Coeficiente de variación	70.6999342	**Media de error estándar**	9.60938488

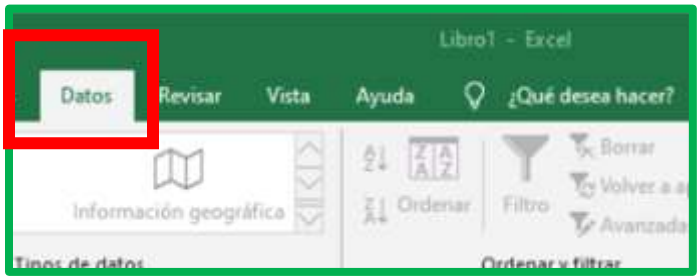
Medidas estadísticas básicas

Localización		Variabilidad	
Media	47.08333	Desviación típica	33.28789
Mediana	45.00000	Varianza	1108
Moda	.	Rango	92.00000

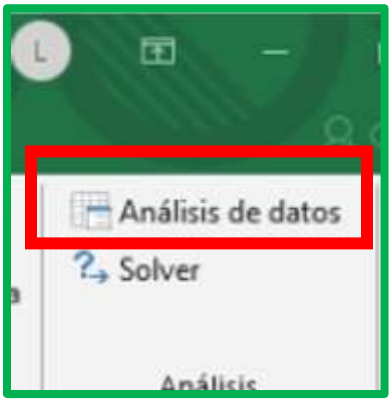
Conclusión: El programa **PROC UNIVARIATE** es de bastante ayuda en el análisis de datos, no necesita ningún comando extra, lo que facilita su uso.

Tema 5. Excel en el análisis de variables de tendencia central y de dispersión

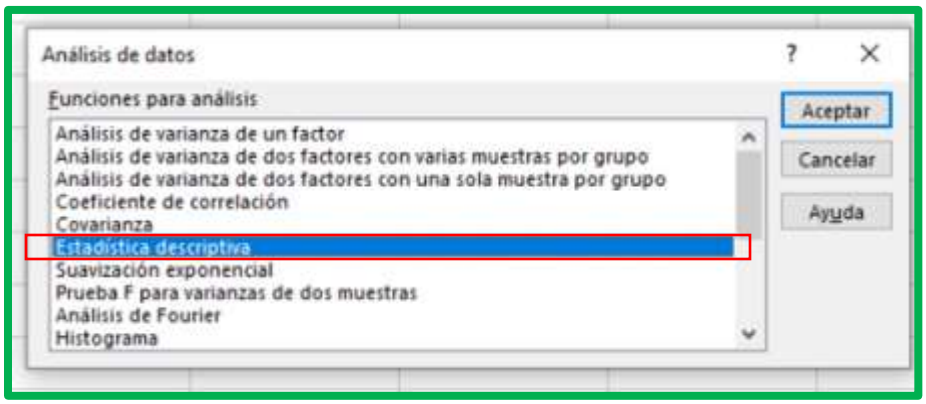
En una hoja de calculo en excel, en la barra de herramientas, en el cuarto lugar despues de archivo se localiza la herramienta datos.



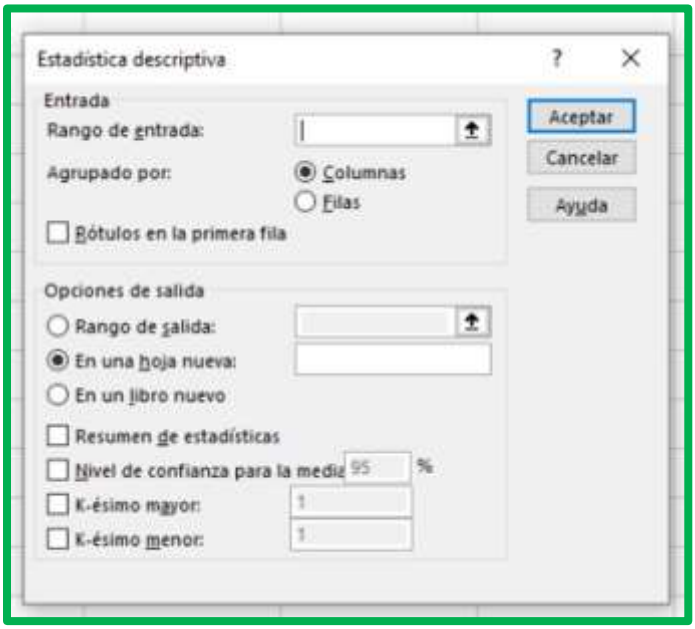
En datos en la parte final se encuentra análisis de datos, una herramienta muy útil en el análisis. En caso de no encontrarse es necesario activarla de la siguiente manera: ir a archivo, en la parte final dar clik en **Opciones**, posteriormente doble clik en **Complementos,** en la pestaña **administrar** seleccionar **Complementos de Excel** y dar clik en **Ir**, se desplegará una ventana en la cual es necesario palomear **Herramientas para Análisis** y dar clik en **aceptar**, posteriormente **Análisis de Datos** se encontrará al final de la pestaña **Datos**.



Estando en análisis de datos se desplegará un menú en el cual es necesario seleccionar "**Estadística descriptiva y dar click en aceptar**".



Al seleccionar **Estadística descriptiva** se desplegará una nueva ventana con una serie de casillas que es necesario llenar con la información del conjunto de datos que se desean analizar.



En el **Rango de entrada** se introduce desde el primer dato hasta el último, se selecciona **columna** debido a que los datos están ordenados en una columna y no en fila, la celda de **Rótulos** se deja bacía, seleccionar que la **Salida** sea **En una hoja nueva** y por último seleccionar **Resumen de estadísticas.**

Ejemplo: Calcula la media, error típico o error estándar, moda, mediana, varianza, desviación estándar, máximo y mínimo de la serie de datos de la **variable X** [ejemplo resuelto en SAS en el Tema 4].

Variable
X
50
40
22
17
99
15
60
7
77
12
97
69

		Columna1
2		
3	Media	47.0833333
4	Error típico	9.60938488
5	Mediana	45
6	Moda	#N/D
7	Desviación estándar	33.2878857
8	Varianza de la muestra	1108.08333
9	Curtosis	-1.35032772
10	Coeficiente de asimetría	0.35437709
11	Rango	92
12	Mínimo	7
13	Máximo	99
14	Suma	565
15	Cuenta	12
16		

Conclusión:

La herramienta **Estadística descriptiva** de Excel es sencilla de utilizar y el programa no tiene costo, debido que se encuentra dentro del paquete de **Excel,** los resultados son similares a los encontrados en el programa SAS en el Tema 4.

Estudio de caso: "Variables de tendencia central y de dispersión en pollos de engorda"

Un granjero crio pollos de engorda de la línea ROSS 308, y los venderá en una feria ganadera, para comprarlos tres jueces analizaran el peso vivo de las aves. El primer juez calculara la media, varianza, desviación estándar, error estándar y coeficiente de variación de acuerdo a las fórmulas. El segundo juez estimara las mismas variables estadísticas utilizando el programa SAS y comparará los resultados del programa con los del primer juez, mientras que, el tercer juez analizara los datos en Excel. Los resultados serán comparados por los tres jueces.

Datos:

Peso vivo en g de 11 pollos de engorda de la línea ROSS 308										
3100	3000	2999	3250	3150	3125	2800	2998	3000	2900	3005

Para el análisis de la varianza y cada uno de sus componentes se utilizará la fórmula:

$$\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]$$

	Peso vivo en g de 11 pollos de engorda de la línea ROSS 308										
Xi	3100	3000	2999	3250	3150	3125	2800	2998	3000	2900	3005
Peso vivo expresado en kg											
Xi	3.1	3	2.999	3.250	3.150	3.125	2.8	2.998	3	2.900	3.005
$Sumatoria\ de\ Xi = \sum_{i=1}^{n} 33327g$ en kg 33.327 *Tamaño de muestra* (n)=11 pollos											

$Promedio = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{33327}{11} = 3029.727\ g$ ó 3.029 kg

El peso vivo de cada pollo fue elevado al cuadrado

Xi^2	9.61	9	8.99	10.56	9.92	9.77	7.84	8.99	9	8.41	9.03

$Suma\ de\ cuadrados\ no\ corregidos = \sum_{i=1}^{n} Xi^2 = 101.123\ kg^2$

$Factor\ de\ corrección = \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n} = \frac{(33.327)^2}{11} = \frac{1110.69}{11} = 100.97\ kg^2$

Suma de cuadrados corregidos $\sum_{i=1}^{n} Xi^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n} = 101.123 - 100.97 = 0.150934\ kg^2$

Varianza:

$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right] = \frac{0.153}{10} = 0.01509342 Kg^2$ y en gramos= $15093.42 g^2$

Desviación estándar:

$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]} = \sqrt{0.0153} = 0.1228$ kg expresado en g=122.8

Error estándar:

$$\frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]}}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{0.1228}{3.3166} = 0.03704\ kg\ y\ en\ gramos = 37.04$$

Coeficiente de variación:

$$\frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]}}{\bar{X}} * 100 = 4.05\%$$

Análisis del peso vivo de los pollos expresado en gramos

Número de pollos	Xi	Xi^2
1	3100	9610000
2	3000	9000000
3	2999	8994001
4	3250	10562500
5	3150	9922500
6	3125	9765625
7	2800	7840000
8	2998	8988004
9	3000	9000000
10	2900	8410000
11	3005	9030025

$$\boldsymbol{Sumatoria\ de\ Xi = \sum_{i=1}^{n} 33327g}$$

n=11

$$\boldsymbol{Promedio = \frac{\sum_{i=1}^{n} Xi}{n} = \frac{33327}{11} = 3029.727\ g}$$

$$\boldsymbol{Suma\ de\ cuadrados\ no\ corregidos = \sum_{i=1}^{n} Xi^2 = 101122655}$$

$$\boldsymbol{Factor\ de\ corrección = \frac{(\sum_{i=1}^{n} Xi)^2}{n}} = 10097172$$

Suma de cuadrados corregidos $\boldsymbol{\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n} = 150934.182}$

$$\boldsymbol{S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right] = 15093.4182}$$

$$S = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]} = 122.8552\ g$$

$$E.E. = \frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]}}{\sqrt{n}} = \frac{S}{\sqrt{n}} = 37.0422\ g$$

$$C.V. = \frac{\sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{1=1}^{n} Xi)^2}{n}\right]}}{\bar{X}} * 100 = 4.05\%$$

En Excel se utilizó la herramienta Análisis de datos y en el catálogo de funciones para el análisis se utilizó la función estadística descriptiva:

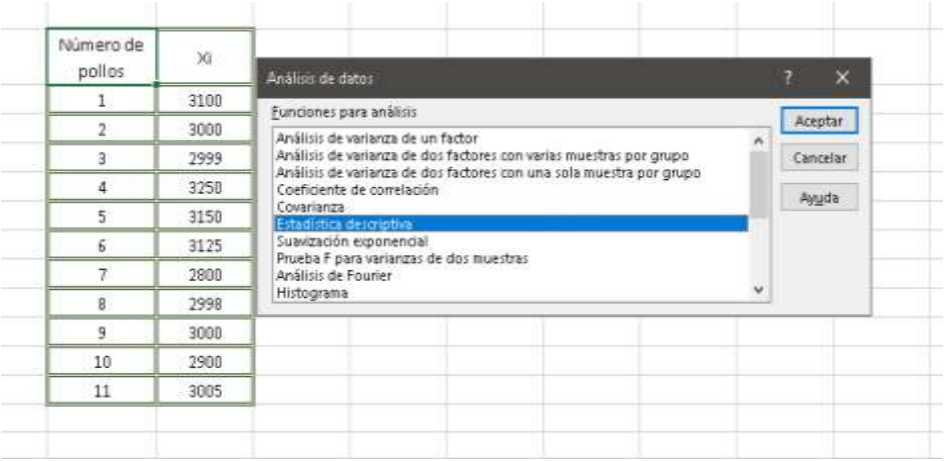


Figura 4. En Excel la función para el análisis es: estadística descriptiva.

Figura 5. En estadística descriptiva.

Columna1	
Media	3029.727273
Error típico o Error Estándar	37.04225865
Mediana	3000
Moda	3000
Desviación estándar	122.8552733
Varianza de la muestra	15093.41818
Mínimo	2800
Máximo	3250
Suma	33327
Cuenta	11

Figura 6. Resultados del análisis estadístico de la variable peso vivo de los once pollos de la línea Ross 308.

Conclusión: **Los resultados en Excel con la función para análisis estadística descriptiva son similares a los encontrados realizando las fórmulas de forma manual.**

Análisis en SAS

En SAS se utilizó el procedimiento PROC UNIVARIATE para realizar el análisis de los 11 datos de peso vivo, este procedimiento realiza un resumen estadístico de los datos:

```
Data Estudio de Caso;
Input PV;
Cards;
3100
3000
2999
3250
3150
3125
2800
2998
3000
2900
3005
;
Proc Univariate;
Var PV;
Run;
```

```
                                Sistema SAS          10:06 Monday, January 25, 2023   5
                                 Procedimiento UNIVARIATE
                                       Variable:  PV

                                          Momentos
   N                                          11    Pesos de la suma                    11
   Media                                3029.72727    Observaciones de la suma         33327
   Desviación típica                    122.855273    Varianza                     15093.4182
  Suma de cuadrados no corregidos       101122655
  Suma de cuadrados corregidos          150934.182
   Coeficiente de variación            4.05499447    Media de error estándar      37.0422586

                                Medidas estadísticas básicas
                          Localización                 Variabilidad
                      Media     3029.727     Desviación típica     122.85527
                      Mediana   3000.000     Varianza                  15093
                      Moda      3000.000     Rango                 450.00000
```

Conclusión: El granjero venderá 33.327 kg de pollo, los once pollos tienen un promedio de peso vivo de 3.029 kg ± 37 g [Error Estándar], la parvada del granjero se encontraba con muy poco coeficiente de variación [4%], por lo tanto, es muy probable la compra inmediata de las aves.

Tema 6. Regresión lineal simple

La regresión lineal simple es útil para encontrar la fuerza o magnitud de cómo se relacionan dos variables (Baeza-Serrato y Vázquez-López, 2014) **una independiente, que se representa con una X,** y otra **dependiente, que se identifica con una Y**.

La regresión **permite estimar el cambio** de la variable dependiente Y por cada unidad de incremento de la variable independiente X. Además, permite hacer una predicción del comportamiento de las variables estudiadas en un determinado punto o momento (Hernández-Lalinde, 2020).

Estructura general del modelo de regresión lineal simple

El modelo de regresión lineal expresa la relación entre dos o más variables aleatorias. Considérese una variable respuesta o dependiente Y, con un regresor o predictor X. La relación lineal entre Y y X queda definida a partir de la siguiente expresión:

$$Y=\beta_0 + \beta_1 X + Ei$$

Componentes del modelo:

Y= Variable dependiente

β0 y β1= Coeficientes de regresión

β0= Ordenada al origen

β1= Pendiente

Ei= Error aleatorio, con media cero y varianza constante

X= Variable independiente

Donde β0 y β1 y son los coeficientes de regresión o parámetros del modelo y Ei es el **componente del error** o perturbación aleatoria (De la Fuente-Fernández *et al.*, 2011).

La ecuación proveerá una aproximación bastante aceptable de la verdadera relación y que las desviaciones del modelo serán recogidas por Ei. Por otro lado, **β1 y β0 son la pendiente y el intercepto de la recta de regresión**, respectivamente (Szretter-Noste, 2017).

Es necesario estimar los coeficientes de regresión, los valores de b0 y b1 se hallan a través de:

$$\boldsymbol{b0 = \overline{Y} - b1\overline{X}}$$

β0= Ordenada al origen.

$$\mathbf{b1 = \frac{\sum_{i=1}^{n}(Xi - \overline{X})\,(Yi - \overline{Y})}{\sum_{i=1}^{n}(Xi - \overline{X})^2}}$$

β1= Pendiente.

Donde $\overline{\mathbf{X}}$ **y** $\overline{\mathbf{Y}}$ son las medias de la variable independiente y dependiente, respectivamente. Al reunir los términos calculados con las ecuaciones $\boldsymbol{b_0}$ y $\boldsymbol{b_1}$ se obtiene el modelo estimado de regresión simple:

$$\boldsymbol{\hat{Y} = b_0 + b_1 X_i}$$

^Indica que esta es la recta de regresión de la muestra.

Recta de regresión

El objetivo de la regresión lineal es explicar el comportamiento de una variable Y, que denominaremos variable explicada (**dependiente o endógena**), a partir de otra variable X, que llamaremos variable explicativa (**independiente o exógena;** Cardona *et al.*, 2013).

La ecuación para una línea recta donde la variable **dependiente** Y está determinada por la variable **independiente** X es:

$$Y = a + bX$$

Donde:

Y = variable dependiente.
a = intersección en Y.
b = pendiente de la recta.
X = variable independiente.

a por lo tanto es el punto donde la recta corta el eje de las Y. Entonces b por lo tanto es una medida de la pendiente de la recta.

El método de los mínimos cuadrados consiste en buscar los **valores de los parámetros a y b** de manera que la suma de los cuadrados de los residuos sea mínima. Esta recta es la recta de regresión por mínimos cuadrados (Montero *et al.*, 2018).

b: coeficiente de represión, aquí expresa la cantidad en la que varía Y cuando X aumenta en una unidad.

Para determinar el coeficiente de regresión **β**, se necesita (Montero *et al.*, 2018):

1. La suma de productos cruzados (PC).

$$PC = \sum_{i=1}^{n} (Xi - \bar{X})\,(Y_i - \bar{Y})$$

2. Calcular b1

$$\mathbf{b1} = \frac{\sum_{i=1}^{n}(\mathbf{Xi}-\overline{\mathbf{X}})(\mathbf{Y_i}-\overline{\mathbf{Y}})}{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{Xi}-\overline{\mathbf{X}})^{\mathbf{2}}}$$

3. Calcular b0

$$\boldsymbol{b0} = \overline{\boldsymbol{Y}} - \boldsymbol{b1}\overline{\boldsymbol{X}}$$

La recta de regresión es la que mejor ajusta a la nube de puntos en el sentido de los mínimos cuadrados (Lina *et al.*, 2006).

La siguiente nomenclatura fue extraída del artículo de Montero *et al.* (2018):

S_{xx}= Suma de cuadrados de X

$$\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{Xi}-\overline{\mathbf{X}})^{\mathbf{2}}$$

S_{yy}= Suma de cuadrados de Y

$$\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{Yi}-\overline{\mathbf{Y}})^{\mathbf{2}}$$

S_{xy}= Suma de Productos Cruzados XY

$$\mathbf{PC} = \sum_{i=1}^{n}(\mathbf{Xi}-\overline{\mathbf{X}})\,(\mathbf{Y_i}-\overline{\mathbf{Y}})$$

Pendiente o coeficiente de regresión: $\mathrm{b1} = \frac{S_{xy}}{S_{xx}}$

Intercepto o coeficiente de regresión: $b_0 = \bar{Y} - \mathrm{b1}\,\bar{\mathrm{x}}$

Ejemplo.

Paso 1. Debes de calcular las medias o promedio de cada grupo.

Datos	
Xi	**Yi**
1	3
2	4
3	7
4	9
5	8
6	2
$\sum_{i=1}^{n} Xi = 21$	$\sum_{i=1}^{n} Yi = 33$
$\bar{X}=3.5$	$\bar{Y}=5.5$

Paso2. Calcular (X_i - $\bar{X}$)

Xi	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
1	(1 – 3.5) = -2.5	6.25
2	(2 - 3.5) = -1.5	2.25
3	(3 - 3.5) = -0.5	0.25
4	(4 - 3.5) = 0.5	0.25
5	(5 - 3.5) = 1.5	2.25
6	(6 - 3.5) = 2.5	6.25
		$\sum_{i=1}^{n}(Xi - \bar{X})^2 = 17.5$

Paso 3. Calcular $(Y - \bar{Y})$

Yi	**$(Y_i - \bar{Y})$**	**$(Y_i - \bar{Y})^2$**
3	(3 – 5.5) = -2.5	6.25
4	(4 – 5.5) = -1.5	2.25
7	(7 – 5.5) = 1.5	2.25
9	(9 – 5.5) = 3.5	12.25
8	(8 – 5.5) = 2.5	6.25
2	(2 – 5.5) = -3.5	12.25
		$\sum_{i=1}^{n}(Yi - \bar{Y})2 = 41.5$

Esta sumatoria es opcional, puesto que no se ocupa en la fórmula.

Paso 4. Calcular $(X_i - \bar{X})(Y_i - \bar{Y})$

$(X_i - \bar{X})$	**$(Y_i - \bar{Y})$**	**$(X_i - \bar{X})(Y_i - \bar{Y})$**
-2.5	-2.5	6.25
-1.5	-1.5	2.25
-0.5	1.5	-0.75
0.5	3.5	1.75
1.5	2.5	3.75
2.5	-3.5	-8.75
		$\sum_{i=1}^{n}(Xi - \bar{X})(Y_i - \bar{Y}) = 4.5$

Paso 5. Calcular el coeficiente de regresión b1 o **pendiente** con la siguiente fórmula:

$$\mathbf{b1} = \frac{\sum_{i=1}^{n}(Xi-\bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Xi-\bar{X})^2}$$

Sustituyendo los valores en la fórmula.

$$\mathbf{b_1} = \frac{4.5}{17.5} = 0.257$$

La pendiente de este ejemplo es de 0.257, sin embargo, es necesario definir que la pendiente: proviene del latín, del verbo "pendere", cuyo significado puede entenderse como "colgar". Matemáticamente la pendiente de una recta se define como la relación del cambio vertical con respecto a la horizontal, también se puede definir con base en el ángulo que forma con respecto al eje X, el punto de referencia para medir la inclinación de una recta sería el eje X.

Paso 6. Usar la siguiente fórmula.

$$\mathbf{b_0 = \overline{Y} - b_1 \overline{X}}$$

Sustituyendo los valores.

$$b_0 = 5.5 - 0.257\ (3.5)$$

$$b_0 = 5.5 - 0.8995$$

$$\mathbf{b_0} = 4.6$$

Paso 7. Sustituyendo los valores en la ecuación de la recta.

$$\hat{Y} = b_0 + b_1 X_i$$

$$\hat{Y} = 4.6 + 0.257X_i$$

Estimando los valores de Yi con los valores de Xi y los coeficientes de regresión b1 y b0 con la ecuación de la recta:

X	b0+b1Xi	Y
1	4.6+0.25 (1) =	6.25
2	4.6+0.25 (2) =	6.5
3	4.6+0.25 (3) =	6.75
4	4.6+0.25 (4) =	7
5	4.6+0.25 (5) =	7.25
6	4.6+0.25 (6) =	7.5

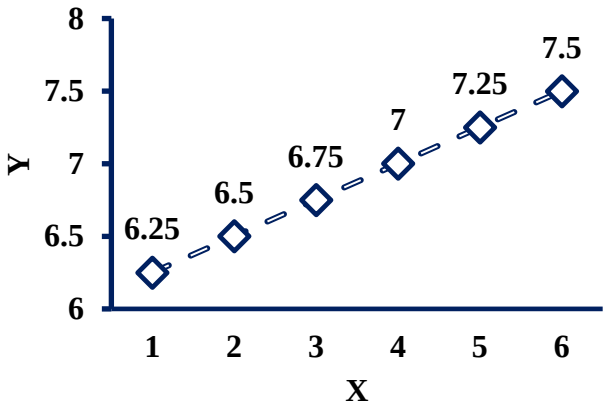


Figura 7. Ecuación de la recta de regresión lineal.

Literatura citada

Montero JAE, Crippa SF, Iogna PGN, Rivas KA. 2018. Modelo de Regresión Lineal Simple para estimar el Peso en función de la Altura de los Estudiantes de la Facultad de Ingeniería de la Universidad Nacional de Jujuy a partir de una muestra. URL: https://fdocuments.es/document/modelo-de-regresin-lineal-simple-para-estimar-el-peso-en-este-trabajo-trata.html?page=6

Baeza-Serrato, Vázquez-López, 2014. Transición de un modelo de regresión lineal múltiple predictivo, a un modelo de regresión no lineal simple explicativo con mejor nivel de predicción: Un enfoque de dinámica de sistemas. Revista Facultad de ingeniería Universidad de Antioquia (71): 59-71.

Cardona MDF, González Rodríguez JL, Rivera Lozano M, Cárdenas Vallejo E. 2013. Inferencia estadística Módulo de regresión lineal simple. Editorial Universidad del Rosario Bogotá D.C. ISSN: 0124-8219. 60 pp.

De la Fuente FS 2011. Regresión múltiple. Documento inédito. Madrid: Universidad Autónoma de Madrid. Recuperado de: URL: http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/REGRE_MULTIPLE /regresion-multiple.pdf

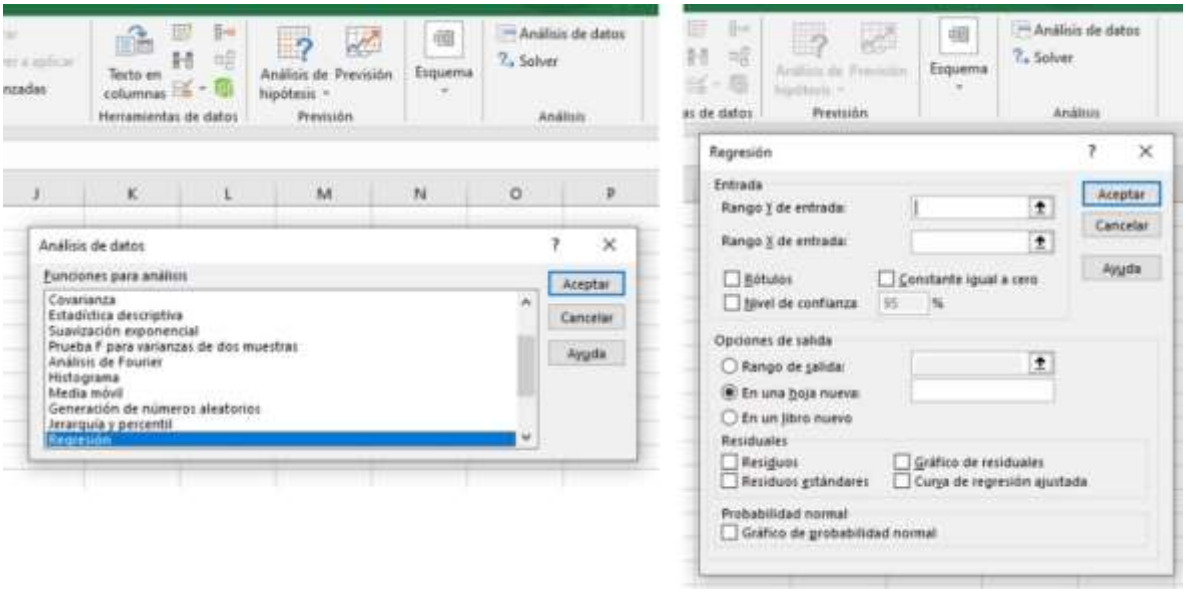
Hernández-Lalinde J, Espinoza-Castro J-F, García-Álvarez D, Bermúdez-Prirela V. 2020. Sobre el uso adecuado de la regresión lineal: conceptualización básica mediante un ejemplo aplicado a las ciencias de la salud. Archivos Venezolanos de Farmacología y Terapéutica.URL: file:///C:/Users/Lenovo/Downloads/Sobreelusoadecuadodelaregresinlineal.pdf

Lina LA, Beatriz ME, Rubio N. 2006. Análisis didáctico de regresión y correlación para la enseñanza media. Revista latinoamericana de investigación en matemática educativa. 9(3): 383-406.

Szretter Noste María Eugenia. 2017. Apunte de Regresión Lineal. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. URL: http://mate.dm.uba.ar/~meszre/apunte_regresion_lineal_szretter.pdf

Tema 7. Regresión lineal simple en Excel

Para realizar una regresión lineal simple en Excel es necesario **abrir una hoja, ubicar la herramienta Datos** (dar click) y en la parte final abrir **análisis de datos**, se despliega un catálogo de funciones para análisis, seleccionar **regresión. En entrada se solicitan los valores de la variable Y y de la variable X.**



En la pestaña de **Rango Y** de entrada solo se deben de introducir los datos de la **variable Y**, posterior mente los datos de la **variable X**, ambos conjuntos de datos deben ser **introducidos sin rótulos**, por tal motivo **no se selecciona** la casilla de Rótulos, en las opciones de salida es preferible seleccionar en una hoja nueva, para que los resultados no se confundan con otros datos.

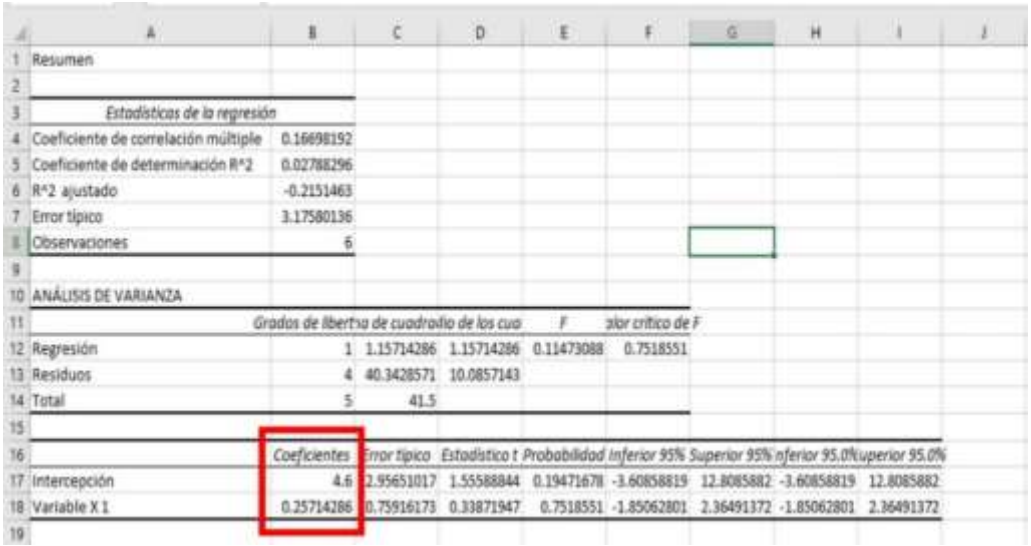
Ejemplo:

El ejemplo realizado anterior mente se va a confirmar utilizando el análisis de datos en Excel.

Datos	
Xi	**Yi**
1	3
2	4
3	7
4	9
5	8
6	2

Resultados del análisis

Los resultados aparecen en una hoja nueva, la intercepción (b0= ordenada al origen), variable X1 (b1=pendiente).



	A	B	C	D	E	F	G	H	I	J
1	Resumen									
2										
3	*Estadísticas de la regresión*									
4	Coeficiente de correlación múltiple	0.16698192								
5	Coeficiente de determinación R^2	0.02788296								
6	R^2 ajustado	-0.2151463								
7	Error típico	3.17580136								
8	Observaciones	6								
9										
10	ANÁLISIS DE VARIANZA									
11		*Grados de libertad*	*de cuadrados*	*de los cuad*	*F*	*Valor crítico de F*				
12	Regresión	1	1.15714286	1.15714286	0.11473088	0.7518551				
13	Residuos	4	40.3428571	10.0857143						
14	Total	5	41.5							
15										
16		*Coeficientes*	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95.0%*	*Superior 95.0%*	
17	Intercepción	4.6	2.95651017	1.55588844	0.19471678	-3.60858819	12.8085882	-3.60858819	12.8085882	
18	Variable X 1	0.25714286	0.75916173	0.33871947	0.7518551	-1.85062801	2.36491372	-1.85062801	2.36491372	
19										

Conclusión:

En Excel se realiza el análisis con mayor rapidez, los valores del ejemplo son similares a los obtenidos en el ejemplo anterior, desarrollando todas las fórmulas.

Tema 8. Regresión lineal simple en SAS

```
Title "REGRESIÓN LINEAL";
Data Reg;
Input X Y;
Cards;
1     3
2     4
3     7
4     9
5     8
6     2

;
Proc reg;
model Y=X;
Run;
```

REGRESION LINEAL 14:46 Thursday, August 5, 2020

Procedimiento REG
Modelo: MODEL1
Variable dependiente: Yte

Parámetros estimados

Variable	DF	Parameter Estimate	Standard Error	Valor t	Pr > |t|
Término i	1	**4.60000**	2.95651	1.56	0.1947
Xdependie	1	**0.25714**	0.75916	0.34	0.7519

Conclusión:

En SAS el análisis se realiza con mayor rapidez, los valores del ejemplo son similares a los obtenidos en el ejemplo anterior, desarrollando todas las fórmulas.

Tema 9. Productos cruzados

El termino productos cruzados (PC) es utilizado en la regresión lineal, es el numerador de la fórmula utilizada para encontrar el valor de b1, también denominado coeficiente de regresión β1. Productos cruzados también es utilizado para calcular el coeficiente de correlación, es el numerador en dicha fórmula.

Fórmula del coeficiente de regresión β1 ó b1:

$$b1 = \frac{\sum_{i1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sum_{i1}^{n}(Xi - \overline{X})^2}$$

Fórmula del coeficiente de correlación:

$$r = \frac{\sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(Xi - \overline{X})^2 \sum_{i=1}^{n}(Yi - \overline{Y})^2}}$$

En ambas fórmulas productos cruzados (PC) es el numerador, sin embargo, para encontrar el valor de PC se encuentran citadas dos fórmulas distintas, la primera fórmula en el libro de Stell y Torrer (1985) y la segunda en el libro de Infante y Zarate (2012).

La primera fórmula citada en el libro Steell y Torrie (1985)

$$PC = \sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})$$

En la segunda fórmula citada en el libro de Infante y Zarate (2012):

$$\boldsymbol{PC = \sum_{i=1}^{n} XiYi - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n}}$$

Productos cruzados (PC) *utilizando las dos fórmulas anteriores*

En el ejemplo se utilizarán los mismos valores de Xi y de Yi que se utilizaron en el ejercicio de regresión lineal [Temas: 6, 7 y 8].

Utilizando la primera fórmula (Steell y Torrie, 1985)

$$PC = \sum_{i=1}^{n}(Xi - \overline{X})(Yi - \overline{Y})$$

Paso 1. Calcular las medias de cada grupo.

Datos	
Xi	**Yi**
1	3
2	4
3	7
4	9
5	8
6	2
$\sum_{i=1}^{n} Xi = 21$	$\sum_{i=1}^{n} = 33$
$\overline{X}$=3.5	$\overline{Y}$=5.5

Paso2. A cada valor de la variable Xi restarle la media ($X_i - \overline{X}$)

Xi	**($X_i - \overline{X}$)**
1	(1 – 3.5) = -2.5
2	(2 - 3.5) = -1.5
3	(3 - 3.5) = -0.5
4	(4 - 3.5) = 0.5
5	(5 - 3.5) = 1.5
6	(6 - 3.5) = 2.5

Paso 3. A cada valor de la variable Yi restarle la media ($Yi - \overline{Y}$)

Yi	**($Y_i - \overline{Y}$)**
3	(3 – 5.5) = -2.5
4	(4 – 5.5) = -1.5
7	(7 – 5.5) = 1.5
9	(9 – 5.5) = 3.5
8	(8 – 5.5) = 2.5
2	(2 – 5.5) = -3.5

Paso 4. Encontrar la sumatoria de productos cruzados $\sum_{i=1}^{n}(\mathbf{Xi} - \overline{\mathbf{X}})$ **($Y_i - \overline{Y}$)**

($X_i - \overline{X}$)	**($Y_i - \overline{Y}$)**	**($X_i - \overline{X}$) ($Y_i - \overline{Y}$)**
-2.5	-2.5	6.25
-1.5	-1.5	2.25
-0.5	1.5	-0.75
0.5	3.5	1.75
1.5	2.5	3.75
2.5	-3.5	-8.75
		$\mathbf{PC = \sum_{i=1}^{n}(Xi - \overline{X})\ (Y_i - \overline{Y}) = 4.5}$

Productos Cruzados:

$$\mathbf{PC = \sum_{i=1}^{n}(Xi - \overline{X})\ (Y_i - \overline{Y}) = 4.5}$$

Utilizando la segunda fórmula publicada en el libro de Infante y Zarate (2012):

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}}$$

Paso 1. Multiplica los valores de Xi por los valores de Yi, el resultado es XiYi.

Datos		
Xi	**Yi**	**XiYi**
1	3	(1)(3) =3
2	4	(2)(4) =8
3	7	(3)(7) =21
4	9	(4)(9) =36
5	8	(5)(8) =40
6	2	(6)(2) =12
$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathrm{Xi} = 21$	$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathrm{Yi} = 33$	$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} = \mathbf{120}$

Paso 2. Multiplica los resultados de la sumatoria de $\sum_{i=1}^{n} \mathrm{Xi}$ por $\sum_{i=1}^{n} \mathrm{Yi}$, y el resultado dividirlo entre n (tamaño de muestra)

$$\frac{(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Xi})(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Yi})}{\mathbf{n}} = \frac{(\mathbf{21})(\mathbf{33})}{\mathbf{6}} = \mathbf{115.5}$$

Paso 3. A la sumatoria $\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi}$ restarle el valor $\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\boldsymbol{n}}$

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}}$$

$$\mathbf{PC = 120 - 115.5 = 4.5}$$

Conclusión: **en ambas fórmulas se obtiene el mismo resultado.**

Tema 10. Suma de cuadrados: $SC = \sum_{i=1}^{n}(Xi - \bar{X})^2$

La suma de cuadrados participa en la varianza de la muestra, es el numerador de la varianza, mientras que, en la regresión lineal participa como denominador; en la fórmula para encontrar b1. **Una propiedad de la media es que la suma de las desviaciones es cero;** $\sum_{i=1}^{n}(Xi - \bar{X}) = 0$**,** si esta propiedad no se cumple la suma de cuadrados será errónea.

La varianza de la muestra se define como la suma de las desviaciones al cuadrado $\sum_{i=1}^{n}(Xi - \bar{X})^2$ dividida entre n-1.

$$S^2 = \frac{\sum_{i=1}^{n}(Xi-\bar{X})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{n}(Xi - \bar{X})^2\right]$$

Para estimar la varianza es necesario desarrollar el factor de corrección y restar lo a la Suma de Cuadrados no Corregidos, de esta manera el resultado obtenido será conocido como Suma de Cuadrados Corregidos. Fórmula citada en el libro de Infante y Zarate (2012):

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Xi^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n}\right]$$

Donde:

$\frac{1}{n-1}$= Uno entre los grados de libertad

$\sum_{i=1}^{n} Xi^2$ = Suma de cuadrados no corregidos

$\frac{(\sum_{i=1}^{n} Xi)^2}{n}$ = Factor de corrección

Al restar el Factor de Corrección a la Suma de Cuadrados no Corregida se obtiene la Suma de Cuadrados Corregidos. **Mientras que, en la primera fórmula si se multiplican los**

grados de libertad por la varianza es posible conocer la Suma de Cuadrados Corregidos, de la siguiente manera:

$$(n-1)S^2 = \sum_{i=1}^{n}(Xi - \overline{X})^2$$

Entonces si no conocemos la Suma de Cuadrados Corregidos podemos estimarlos sabiendo el valor de la desviación estándar y los grados de libertad. El numerador de S^2 se conoce como la Suma de Cuadrados y a menudo se denota SC. La fórmula de definición de la Suma de Cuadrados se reduce a la siguiente fórmula de trabajo para los cálculos:

$$\sum_{i=1}^{n}(Xi - \overline{X})^2 = \left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}\right]$$

En el primer término los valores se elevan al cuadrado y luego se suman, en el segundo término se resta el Factor de Corrección a la Suma de Cuadrados no Corregidos.

Ejerció para confirmar que con las dos fórmulas se obtiene el mismo valor

Fórmula 1:

$$\sum_{i=1}^{n}(Xi - \overline{X})^2$$

Estimar la suma de cuadrados de Xi con la fórmula $\sum_{i=1}^{n}(\mathrm{Xi} - \overline{\mathrm{X}})^2$

Xi	**$(X_i - \overline{X})$**	**$(X_i - \overline{X})^2$**
1	(1 – 3.5) = -2.5	6.25
2	(2 - 3.5) = -1.5	2.25
3	(3 - 3.5) = -0.5	0.25
4	(4 - 3.5) = 0.5	0.25
5	(5 - 3.5) = 1.5	2.25
6	(6 - 3.5) = 2.5	6.25
n=6	$\sum_{i=1}^{n}(\mathbf{Xi} - \overline{\mathbf{X}}) = \mathbf{0}$	$\sum_{i=1}^{n}(\mathbf{Xi} - \overline{\mathbf{X}})^2 = \mathbf{17.5}$

Conclusión:

Suma de Cuadrados Corregidos con la fórmula $\sum_{i=1}^{n}(\mathrm{Xi} - \overline{\mathrm{X}})^2 = 17.5$

Fórmula 2:

Estimar la suma de cuadrados de Xi con la formula $\sum_{i=1}^{n} \mathrm{Xi}^2 - \frac{(\sum_{i=1}^{n} \mathrm{Xi})^2}{n}$

Xi	**Xi²**
1	$(1)^2 = 1$
2	$(2)^2 = 4$
3	$(3)^2 = 9$
4	$(4)^2 = 16$
5	$(5)^2 = 25$
6	$(6)^2 = 36$
$\sum_{i=1}^{n} \mathrm{Xi} = 21$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 91$
$\frac{(\sum_{i=1}^{n} Xi)^2}{n} = \frac{(21)^2}{6} = 73.5$	$SC = \sum_{i=1}^{n} \mathrm{Xi}^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n} = 91 - 73.5 = \mathbf{17.5}$

Paso 1. Cada valor de Xi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^2$.

Paso 2. La sumatoria de Xi $[\sum_{i=1}^{n} Xi]$ elévala al cuadrado y divídela entre n [n= tamaño de la muestra]; $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$

Paso 3. Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^2$ **menos** $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$

Conclusión:

Suma de Cuadrados Corregidos con la fórmula $= \sum_{i=1}^{n} \mathrm{Xi}^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n} =$ 91- 73.5 = 17.5. *El resultado encontrado en ambas formulas es similar.*

Estudio de caso: Regresión lineal de las variables: kilogramos de queso y litros de leche

Instrucciones: estima los coeficientes de regresión lineal (b1 y b0), el valor de productos cruzados (PC) con las dos fórmulas [Tema 9], el valor de suma de cuadrados (SC) con las tres fórmulas [Tema 10] y diseña la gráfica de regresión con su ecuación.

Valores de Xi (kg de queso) y Yi (litros de leche)

Xi	1	2	3	4	5	6	7
Yi	10	21	33	40	55	63	77

Estimar el valor de productos cruzados [PC] con las dos fórmulas:

Fórmula 1:

$$\mathbf{PC = \sum_{i=1}^{n}(Xi - \bar{X})(Yi - \bar{Y})}$$

Paso 1: Calcular la sumatoria y el promedio de Xi y Yi.

Xi	**Yi**
1	10
2	21
3	33
4	42
5	55
6	63
7	77
$\sum_{i=1}^{n} Xi = 28$	$\sum_{i=1}^{n} \mathrm{Yi} = 301$

$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} = 4$	$\overline{Y} = \frac{\sum_{i=1}^{n} Yi}{n} = 43$

Paso 2: Resta a cada valor de Xi y Yi su media respectivamente

Resta $(Xi - \overline{X})$

Xi	$(\mathbf{Xi} - \overline{\mathbf{X}})$
1	(1 - 4) = -3
2	(2 - 4) = -2
3	(3 - 4) = -1
4	(4 - 4) = 0
5	(5 - 4) = 1
6	(6 - 4) = 2
7	(7 - 4) = 3
$\sum_{i=1}^{n} Xi = 28$	$\sum_{i=1}^{n} (Xi - \overline{X}) = 0$

Resta $(Yi - \overline{Y})$

Yi	$(\mathbf{Yi} - \overline{\mathbf{Y}})$
10	(10 - 43) = -33
21	(21 - 43) = -22
33	(33 - 43) = -10
42	(42 - 43) = -1
55	(55 - 43) = 12
63	(63 - 43) = 20
77	(77 - 43) = 34
$\sum_{i=1}^{n} Yi = 301$	$\sum_{i=1}^{n} (Yi - \overline{Y}) = 0$

Paso 3: Multiplica el resultado de la resta de $(\mathrm{Xi} - \overline{\mathrm{X}})$ por el resultado de la resta de $(\mathrm{Yi} - \overline{\mathrm{Y}})$.

$(\mathbf{Xi} - \overline{\mathbf{X}})$	$(\mathbf{Yi} - \overline{\mathbf{Y}})$	$(\mathbf{Xi} - \overline{\mathbf{X}})\ (\mathbf{Yi} - \overline{\mathbf{Y}})$
(1 - 4) = -3	(10 - 43) = -33	(-3) (-33) = 99
(2 - 4) = -2	(21 - 43) = -22	(-2) (-22) = 44
(3 - 4) = -1	(33 - 43) = -10	(-1) (-10) = 10
(4 - 4) = 0	(42 - 43) = -1	(0) (-1) = 0
(5 - 4) = 1	(55 - 43) = 12	(1) (12) = 12
(6 - 4) = 2	(63 - 43) = 20	(2) (20) = 40
(7 - 4) = 3	(77 - 43) = 34	(3) (34) = 102
		$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{Xi} - \overline{\mathbf{X}})\ (\mathbf{Yi} - \overline{\mathbf{Y}}) = \mathbf{307}$

Fórmula 2:

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}}$$

Paso 1: Multiplica los valores de Xi por los valores de Yi

Xi	**Yi**	**XiYi**
1	10	10
2	21	42
3	33	99
4	42	168

5	55	275
6	63	378
7	77	539
$\sum_{i=1}^{n} Xi = 28$	$\sum_{i=1}^{n} Yi = 301$	$\sum_{i=1}^{n} XiYi = 1511$
$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} = 4$	$\overline{Y} = \frac{\sum_{i=1}^{n} Yi}{n} = 43$	

Paso 2: Multiplica las sumatoria $\sum_{i=1}^{n} Xi$ por la sumatoria $\sum_{i=1}^{n} Yi$ y el resultado de la multiplicación dividirlo entre el tamaño de muestra; $\frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n}$

$$\frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n} = \frac{(28)(301)}{7} = \frac{8428}{7} = 1204$$

Paso 3: Calcula el valor de productos cruzado con la fórmula:

$$PC = \sum_{i=1}^{n} XiYi - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n} = 1511 - 1204 = 307$$

Estima los valores de la Suma de Cuadrados Corregidos con las tres fórmulas siguientes:

Fórmula 1:

$$\sum_{i=1}^{n} (Xi - \overline{X})^2$$

Paso 1: Realiza la sumatoria de Xi y calcula su media

Paso 2: Resta a cada valor de Xi su media; $(Xi - \bar{X})$

Paso 3: Eleva al cuadrado el resultado de la resta suma los valores; $\sum_{i=1}^{n} (Xi - \overline{X})^2$

Xi	$(\mathbf{Xi} - \bar{\mathbf{X}})$	$(\mathbf{Xi} - \bar{\mathbf{X}})^{\mathbf{2}}$
1	(1 - 4) = -3	9
2	(2 - 4) = -2	4
3	(3 - 4) = -1	1
4	(4 - 4) = 0	0
5	(5 - 4) = 1	1
6	(6 - 4) = 2	4
7	(7 - 4) = 3	9
$\sum_{i=1}^{n} \mathrm{Xi} = 28$	$\sum_{i=1}^{n} (\mathrm{Xi} - \bar{\mathrm{X}}) = 0$	$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{(Xi - \bar{X})^2 = 28}$

Fórmula 2:

$$\sum_{i=1}^{n} \mathrm{Xi}^{2} - \frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$$

Paso 1. Cada valor de Xi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^{2}$.

Paso 2. Calcular la sumatoria de Xi $\left[\sum_{i=1}^{n} \mathrm{Xi}\right]$ elévala al cuadrado y divídela entre n [n= tamaño de la muestra]; $\frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$

Paso 3. Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^{2}$ **menos** $\frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$

Xi	$\boldsymbol{Xi^2}$
1	1
2	4
3	9

4	16
5	25
6	36
7	49
$\sum_{i=1}^{n} Xi = 28$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 140$
$\sum_{i=1}^{n} \mathbf{Xi^2}\text{-} \frac{(\sum_{i=1}^{n} Xi)^2}{n} = 140 - \frac{(28)^2}{7} = 140 - 112 = 28$	

Fórmula 3:

$$\left[\sum_{i=1}^{n} Xi^2 - n\bar{X}^2\right]$$

Paso 1. Cada valor de Xi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^2$

Paso 2: Multiplica el tamaño de muestra (n) por la media elevada al cuadrado $\bar{X}^2$

Paso 3: Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^2$ **menos** $n\bar{X}^2$

Xi	Xi^2
1	1
2	4
3	9
4	16
5	25
6	36

7	49
$\sum_{i=1}^{n} Xi = 28$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 140$
$\sum_{i=1}^{n} \mathrm{Xi}^2 - n\bar{X}^2 = 140 - 7(4^2) =$ **140-112= 28**	

Calcula b1 con las tres fórmulas anteriores

Fórmula1:

$$\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{Xi}-\bar{\mathbf{X}})(\mathbf{Yi}-\bar{\mathbf{Y}})}{\sum_{i=1}^{n}(Xi-\bar{X})^2} = \frac{\mathbf{307}}{\mathbf{28}} = \mathbf{10.96}$$

Fórmula 2:

$$\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{XiYi} - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{\mathbf{n}}}{\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}} = \frac{\mathbf{307}}{\mathbf{28}} = \mathbf{10.96}$$

Fórmula 3:

$$\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{XiYi} - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{\mathbf{n}}}{\sum_{i=1}^{n} Xi^2 - n\bar{X}^2} = \frac{\mathbf{307}}{\mathbf{28}} = \mathbf{10.96}$$

Calcula b0:

Fórmula:

$$b0 = \bar{Y} - b1\bar{X}$$

$$b0 = 43 - 10.96(4) = 43 - 43.857 = -0.857$$

Calculando los valores de Yi con distintos valores de Xi, con la recta de regresión es posible estimar cuantos litros de leche se necesitan para obtener los kilogramos de queso que un productor desee producir:

$$\widehat{Y} = b0 + b1Xi$$

Donde: Xi son kilogramos de queso y Yi los litros de leche.

Xi	*bo+b1Xi*	*Yi*
9	*-0.0857+10.96(9)*	*98.55*
13	*-0.0857+10.96(13)*	*142.39*
15	*-0.0857+10.96(15)*	*164.31*
17	*-0.0857+10.96(17)*	*186.23*
21	*-0.0857+10.96(21)*	*230.07*
27	*-0.0857+10.96(27)*	*295.83*
30	*-0.0857+10.96(30)*	*328.71*

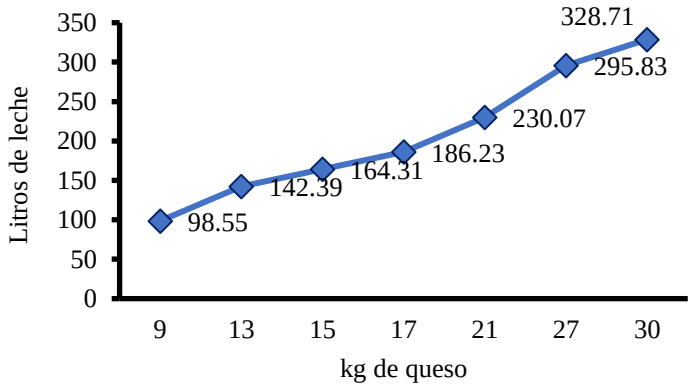


Figura 8. Con la regresión lineal se puede estimar cuantos litros de leche se necesitan para obtener cualquier cantidad de queso que se desee producir.

Regresión en SAS: litros de leche [Xi] *necesarios para producir queso* [Yi]

Encontrar coeficientes de regresión lineal (b1 y b0) con los v***alores de Xi (kg de queso) y Yi (litros de leche)***

```
Data Reg;
Input X Y;
Cards;
1      10
2      21
3      33
4      42
5      55
6      63
7      77
;
Proc Reg;
Model y=x;
Run;
```

```
                          Sistema SAS          10:56 Sunday, February 7, 2023

                               Procedimiento REG
                                Modelo: MODEL1
                          Variable dependiente: Yte

                             Analysis of Variance
                             Parámetros estimados
                          Parameter       Standard
Variable        DF         Estimate          Error    Valor t    Pr > |t|
Término i        1         -0.85714        1.06666      -0.80      0.4581
Xdependie        1         10.96429        0.23851      45.97      <.0001
```

Estudio de caso: Regresión lineal de las variables: borregos en barbacoa y número de platillos

Instrucciones: *Un productor de borregos para barbacoa desea conocer el número de platillos que se producirán de 30 borregos hechos en barbacoa. Para realizar la predicción con anterioridad registró los siguientes datos: Xi: número de borregos en barbacoa, Yi: número de platillos.*

Xi	1	2	3	4	5	
Yi	27	53	77	89	105	

Estimar productos cruzados [PC] con la fórmula:

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{\mathbf{n}}$$

Paso 1: Multiplica los valores de Xi por los valores de Yi

Xi	**Yi**	**XiYi**
1	27	27
2	53	106
3	77	231
4	89	356
5	105	525
$\sum_{i=1}^{n} Xi = 15$	$\sum_{i=1}^{n} Yi = 351$	$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} = \mathbf{1245}$
$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} = 3$	$\overline{Y} = \frac{\sum_{i=1}^{n} Yi}{n} = 70.2$	

Paso 2: Multiplica las sumatoria $\sum_{i=1}^{n} Xi$ por la sumatoria $\sum_{i=1}^{n} Yi$ y la multiplicación dividirla entre el tamaño de muestra; $\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}}$

$$\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}} = \frac{(15)(351)}{5} = \frac{5265}{5} = 1053$$

Paso 3: Calcula el valor de productos cruzado con la fórmula:

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}} = 1245 - 1053 = 192$$

Estimar la Suma de Cuadrados corregidos [SC] con la fórmula:

Fórmula 1:

$$\sum_{i=1}^{n} \mathrm{Xi}^{2} - \frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$$

Paso 1. Elevar cada valor de Xi al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^{2}$.

Paso 2. La sumatoria de Xi $\left[\sum_{i=1}^{n} Xi\right]$ elévala al cuadrado y divídela entre n [n= tamaño de la muestra]; $\frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$

Paso 3. Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^{2}$ **menos** $\frac{\left(\sum_{i=1}^{n} Xi\right)^{2}}{n}$

Xi	Xi^2
1	1
2	4
3	9
4	16
5	25
$\sum_{i=1}^{n} Xi = 15$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 55$
$\sum_{i=1}^{n} \mathrm{Xi}^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n} = 55 - \frac{(15)^2}{5} = 55 - 45 = 10$	

La Suma de Cuadrados corregidos es: $\sum_{i=1}^{n} \mathbf{Xi}^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n} = 55 -$

$$\frac{(15)^2}{5} = 55 - 45 = 10$$

Calcula b1 con la fórmula:

Fórmula:

$$b1 = \frac{\sum_{i=1}^{n} \mathrm{XiYi} - \frac{\left(\sum_{i=1}^{n} Xi\right)\left(\sum_{i=1}^{n} Yi\right)}{n}}{\sum_{i=1}^{n} Xi^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n}} = \frac{192}{10} = 19.2$$

Calcula b0:

Fórmula:

$$b0 = \bar{Y} - b1\bar{X}$$

$$b0 = 70.2 - 19.2(3) = 70.2 - 57.6 = 12.6$$

Con la recta de regresión es posible estimar cuantos platillos se producirán con 30 borregos en barbacoca:

$$\hat{Y} = b0 + b1Xi$$

Xi	*bo+b1Xi*	*Yi*
30	*12.6 + 19.2 (30)*	*588.6*
43	*12.6 + 19.2 (43)*	*838.2*
55	*12.6 + 19.2 (55)*	*1068.6*

Tema 11. Correlación (r) lineal de Pearson

La finalidad de la correlación es examinar la dirección y la fuerza de la asociación entre dos variables (Dagnino, 2014). El coeficiente de correlación es un valor cuantitativo, sus valores de esta oscilan entre -1 y 1. **Si r = 1 se habla de una correlación positiva perfecta**. **Cuando el coeficiente es igual a cero (r = 0) se dice que las variables están incorrectamente relacionadas [no existe relación entre ellas].** La correlación de Pearson perfecta se encuentra entre +1 y -1, en el primer caso la relación es perfecta positiva y en el segundo perfecta negativa (Lahura, 2003).

Para interpretar lo que es el coeficiente de correlación se utilizara la siguiente escala:

Valor	Significado
-1	Correlación negativa grande y perfecta.
-0.9 a -0.99	Correlación negativa muy alta.
-0.7 a -0.89	Correlación negativa alta.
-0.4 a 0.69	Correlación negativa moderada.
-0.2 a -0.39	Correlación negativa baja.
-0.1 a -0.19	Correlación negativa muy baja.
0	Correlación nula.
0.1 a 0.19	Correlación positiva muy baja.
0.2 a 0.39	Correlación positiva baja.
0.4 a 0.69	Correlación positiva moderada.
0.7 a 0.89	Correlación positiva alta.

0.9 a 0.99	Correlación positiva muy alta.
1	Correlación positiva grande y perfecta.

Coeficiente de correlación de Pearson (r) se puede calcular de acuerdo con dos fórmulas, la primera citada en el libro de Steel y Torrie (1985) y la segunda citada en el libro de Infante y Zarate (2012).

La primera fórmula es la siguiente:

$$r=\frac{\sum_{i=1}^{n}(Xi-\bar{x})(Yi-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(Xi-\bar{X})^2\sum_{i=1}^{n}(Yi-\bar{Y})^2}}$$

Donde:

r = coeficiente de correlación.

$\sum_{i=1}^{n}(\mathbf{Xi}-\bar{\mathbf{X}})\,(\mathbf{Y_i}-\bar{\mathbf{Y}})$**;** sumatoria de productos cruzados**.**

$\sum_{i=1}^{n}(\mathbf{Xi}-\bar{\mathbf{X}})^2$**;** suma de cuadrados de la variable Xi.

$\sum_{i=1}^{n}(\mathbf{Yi}-\bar{\mathbf{Y}})^2$**;** suma de cuadrados de la variable Yi.

Pasos para calcular el coeficiente de correlación en la fórmula citada en el libro de Steel y Torrie (1985):

Paso 1. Calcular la media de la variable Xi y de la variable Yi.

Paso 2. A cada valor de la variable Xi restarle la media $\sum_{i=1}^{n}(Xi-\bar{X})$

Paso 3. A cada valor de la variable Yi restarle la media $\sum_{i=1}^{n}(Yi-\bar{Y})$**.**

Paso 4. A cada valor generado en la resta del grupo de datos de la variable Xi multiplicarlo por el valor generado en la resta del grupo Yi, a esta multiplicación se le conoce como productos cruzados $\sum_{i=1}^{n}(\mathbf{Xi}-\overline{\mathbf{X}})\,(\mathbf{Y_i}-\overline{\mathbf{Y}})$ y este es el numerador de la fórmula.

Paso 5. Calcular la suma de cuadrados de Xi; $\sum_{i=1}^{n}(\mathbf{Xi}-\overline{\mathbf{X}})^2$

Paso 6. Calcular la suma de cuadrados de Yi; $\sum_{i=1}^{n}(\mathbf{Yi}-\overline{\mathbf{Y}})^2$

Paso 7. Multiplicar la suma de cuadrados y al resultado estimar su raíz cuadrada

$$\sqrt{\sum_{i=1}^{n}(Xi-\overline{X})^2\sum_{i=1}^{n}(Yi-\overline{Y})^2}$$

Paso 8. Con los resultados anteriores estimar el coeficiente de correlación de acuerdo a la fórmula:

$$r=\frac{\sum_{i=1}^{n}(Xi-\overline{x})(Yi-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(Xi-\overline{X})^2\sum_{i=1}^{n}(Yi-\overline{Y})^2}}$$

Ejemplo. Utilizando los mismos datos que en el ejemplo de regresión lineal

Paso 1. Calcular las medias de cada grupo.

Datos	
Xi	Yi
1	3
2	4
3	7
4	9
5	8
6	2
$\overline{\mathbf{X}}$**=3.5**	$\overline{\mathbf{Y}}$**=5.5**

Paso2. A cada valor de la variable Xi restarle la media ($X_i - \overline{X}$)

Xi	**($X_i - \overline{X}$)**
1	(1 – 3.5) = -2.5
2	(2 - 3.5) = -1.5
3	(3 - 3.5) = -0.5
4	(4 - 3.5) = 0.5
5	(5 - 3.5) = 1.5
6	(6 - 3.5) = 2.5

Paso 3. A cada valor de la variable Xi restarle la media ($Y - \overline{Y}$)

Yi	**($Y_i - \overline{Y}$)**
3	(3 – 5.5) = -2.5
4	(4 – 5.5) = -1.5
7	(7 – 5.5) = 1.5
9	(9 – 5.5) = 3.5
8	(8 – 5.5) = 2.5
2	(2 – 5.5) = -3.5

Paso 4. Encontrar la sumatoria productos cruzados $\sum_{i=1}^{n}(\mathbf{Xi} - \overline{\mathbf{X}})$ **(Y_i -** $\overline{\mathbf{Y}}$

(X_i - $\overline{\mathbf{X}}$**)**	**(Y_i -** $\overline{\mathbf{Y}}$**)**	**(X_i -** $\overline{\mathbf{X}}$**) (Y_i -** $\overline{\mathbf{Y}}$**)**
-2.5	-2.5	6.25
-1.5	-1.5	2.25
-0.5	1.5	-0.75
0.5	3.5	1.75
1.5	2.5	3.75
2.5	-3.5	-8.75
		$\sum_{i=1}^{n}(\mathbf{Xi} - \overline{\mathbf{X}})$ **(Y_i -** $\overline{\mathbf{Y}}$**) = 4.5**

Paso 5 y 6. Calcular la suma de cuadrados de los valores de Xi; $\sum_{i=1}^{n}(\mathbf{Xi} - \bar{\mathbf{X}})^2$ y la suma de cuadrados de los valores de Yi; $\sum_{i=1}^{n}(\mathbf{Yi} - \bar{\mathbf{Y}})^2$

$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
6.25	6.25
2.25	2.25
0.25	2.25
0.25	12.25
2.25	6.25
6.25	12.25
$\sum_{i=1}^{n}(\mathbf{Xi} - \bar{\mathbf{X}})^2$=17.5	$\sum_{i=1}^{n}(\mathbf{Yi} - \bar{\mathbf{Y}})^2$= 41.5

Paso 7. Multiplicar la suma de cuadrados **de Xi y Yi,** al resultado aplicarle raíz cuadrada

$$\sqrt{\sum_{i=1}^{n}(\boldsymbol{Xi} - \bar{\boldsymbol{X}})^2\,(\boldsymbol{Yi} - \bar{\boldsymbol{Y}})^2}$$

Sustituyendo los resultados:

$\sqrt{(17.5)(41.5)} = \sqrt{726.25} = 26.95$

Paso 8. Calcular el coeficiente de correlación $r = \frac{4.5}{26.95} = 0.17$

Mismo ejemplo con la fórmula publicada en el libro de Infante y Zarate (2012):

$$r = \frac{\sum_{i=1}^{n} XiYi - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n}}{\sqrt{\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}\right]\left[\sum_{i=1}^{n} Yi^2 - \frac{(\sum_{i=1}^{n} Yi)^2}{n}\right]}}$$

También escrita de la siguiente manera:

$$\frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Xi})(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Yi})}{\mathbf{n}}}{\left\{\left[\sum_{\mathrm{i=1}}^{\mathrm{n}} Xi^2 - \frac{(\sum_{i=1}^{n} \mathrm{Xi})^2}{\mathrm{n}}\right]\left[\sum_{\mathbf{i=1}}^{\mathbf{n}} \boldsymbol{Yi}^{\mathbf{2}} - \frac{(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \mathbf{Yi})^{\mathbf{2}}}{\mathbf{n}}\right]\right\}^{\frac{1}{2}}}$$

Donde:

r= coeficiente de correlación.

$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Xi})(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Yi})}{\mathbf{n}} =$ sumatoria de productos cruzados**.**

$\sum_{\mathrm{i=1}}^{\mathrm{n}} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{\mathrm{n}}$ = suma de cuadrados de la variable Xi.

$\sum_{\mathbf{i=1}}^{\mathbf{n}} \boldsymbol{Yi}^{\mathbf{2}} - \frac{(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Yi})^{\mathbf{2}}}{\mathbf{n}}$ = suma de cuadrados de la variable Yi.

Pasos para calcular el coeficiente de regresión de acuerdo a la siguiente fórmula:

$$\frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{XiYi} - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{\mathrm{n}}}{\left\{\left[\sum_{\mathrm{i=1}}^{\mathrm{n}} Xi^2 - \frac{(\sum_{i=1}^{n} \mathrm{Xi})^2}{\mathrm{n}}\right]\left[\sum_{\mathrm{i=1}}^{\mathrm{n}} Yi^2 - \frac{(\sum_{i=1}^{n} \mathrm{Yi})^2}{\mathrm{n}}\right]\right\}^{\frac{1}{2}}}$$

Paso 1. Multiplica los valores de Xi por los valores de Yi, el resultado es XiYi

Paso 2. Multiplica los resultados de la sumatoria de $\sum_{i=1}^{n} \mathrm{Xi}$ y $\sum_{i=1}^{n} \mathrm{Yi}$, y el resultado divídelo entre n (tamaño de muestra).

Paso 3. A la sumatoria $\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{XiYi}$ restarle el valor $\frac{(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Xi})(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Yi})}{\mathbf{n}}$

Nota: En el **Paso 3** se ha calculado **el valor de productos cruzados**, los pasos que a continuación se describen **son para calcular los valores de la suma de cuadrados** de las variables Xi y Yi.

Paso 4. Elevar al cuadrado cada valor de Xi y realizar la sumatoria de esos valores; $\sum_{i=1}^{n} Xi^2$

Paso 5. Elevar al cuadrado la sumatoria de Xi $[(\sum_{i=1}^{n} Xi)^2]$, al resultado dividirlo entre el tamaño de muestra (n): $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$.

Paso 6. A la sumatoria $\sum_{i=1}^{n} Xi^2$ restarle el resultado de la división $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$.

Nota: En el **Paso 6** se ha calculado la suma de cuadrados de la variable Xi, los pasos que continúan son para calcular la suma de cuadrados de la variable Yi.

Paso 7. Eleva al cuadrado cada valor de Yi, y realiza la sumatoria de estos valores; $\sum_{i=1}^{n} Yi^2$

Paso 8. Eleva al cuadrado la sumatoria de Yi $[(\sum_{i=1}^{n} Yi)^2]$, al resultado es necesario dividirlo entre el tamaño de muestra (n): $\frac{(\sum_{i=1}^{n} Yi)^2}{n}$.

Paso 9. A la sumatoria $\sum_{i=1}^{n} Yi^2$ restarle el resultado de la división $\frac{(\sum_{i=1}^{n} Yi)^2}{n}$.

Paso 10. Multiplicar la suma de cuadrados de la variable Xi y Yi y al resultado de la multiplicación elevarlo al exponente un medio o aplicarle raíz cuadrada:

$$\left\{\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}\right]\left[\sum_{i=1}^{n} Yi^2 - \frac{(\sum_{i=1}^{n} Yi)^2}{n}\right]\right\}^{\frac{1}{2}}$$

$$\sqrt{\left[\sum_{i=1}^{n} Xi - \frac{(\sum Xi)^2}{n}\right]\left[\sum_{i=1}^{n} Yi - \frac{(\sum Yi)^2}{n}\right]}$$

Paso 11. el valor de productos cruzados dividirlo entre el resultado del exponente a un medio o de la raíz cuadrada, este valor es el coeficiente de regresión:

$$\mathbf{r} = \frac{\sum_{i=1}^{n} \mathbf{XiYi} - \frac{(\sum \mathbf{Xi})(\sum \mathbf{Yi})}{\mathbf{n}}}{\left\{\left[\sum_{i=1}^{n} Xi - \frac{(\sum Xi)^2}{n}\right]\left[\sum_{i=1}^{n} \mathbf{Yi} - \frac{(\sum \mathbf{Yi})^2}{\mathbf{n}}\right]\right\}^{\frac{1}{2}}}$$

Ejemplo. Se utilizarán los datos del ejemplo de regresión lineal para ser analizados con la fórmula:

$$\frac{\sum_{i=1}^{n} \mathbf{XiYi} - \frac{(\sum \mathbf{Xi})(\sum \mathbf{Yi})}{\mathbf{n}}}{\left\{\left[\sum_{i=1}^{n} Xi - \frac{(\sum Xi)^2}{n}\right]\left[\sum_{i=1}^{n} \mathbf{Yi} - \frac{(\sum \mathbf{Yi})^2}{\mathbf{n}}\right]\right\}^{\frac{1}{2}}}$$

Datos de Xi y Yi:

Datos	
Xi	**Yi**
1	3
2	4
3	7
4	9
5	8
6	2

Paso 1. Multiplica los valores de Xi por los valores de Yi, el resultado es XiYi

Datos		
Xi	**Yi**	**XiYi**
1	3	(1)(3) =3
2	4	(2)(4) =8
3	7	(3)(7) =21
4	9	(4)(9) =36
5	8	(5)(8) =40
6	2	(6)(2) =12
$\sum_{i=1}^{n} Xi = 21$	$\sum_{i=1}^{n} Yi = 33$	$\sum_{i=1}^{n} XiYi = 120$

Paso 2. Multiplica los resultados de la sumatoria de $\sum_{i=1}^{n}$ Xi y $\sum_{i=1}^{n}$ Yi, y el resultado divídelo entre n (tamaño de muestra)

$$\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\boldsymbol{n}} = \frac{(21)(33)}{6} = \mathbf{115.5}$$

Paso 3. A la sumatoria $\sum_{i=1}^{n}$ XiYi restarle el valor $\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\boldsymbol{n}}$

El resultado de productos cruzados (PC) es el siguiente:

$$\mathbf{PC = 120 - 115.5 = 4.5}$$

Paso 4. Eleva al cuadrado cada valor de Xi y realiza la sumatoria de esos valores; $\sum_{i=1}^{n} Xi^2$

Datos	
Xi	**Xi^2**
1	**$(1)^2 = 1$**
2	**$(2)^2 = 4$**
3	**$(3)^2 = 9$**
4	**$(4)^2 = 16$**
5	**$(5)^2 = 25$**
6	**$(6)^2 = 36$**
$\sum_{i=1}^{n} Xi = 21$	$\sum_{i=1}^{n} Xi^2 = 91$
$\sum_{i=1}^{n} \frac{(Xi)^2}{n} = \frac{(21)^2}{6} = \frac{441}{6} = 73.5$	

Paso 5. Eleva al cuadrado la sumatoria de Xi, el resultado dividirlo entre el tamaño de muestra (n); $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$

Paso 6. A la sumatoria $\sum_{i=1}^{n} Xi^2$ restarle el resultado de la división $\frac{(\sum_{i=1}^{n} Xi)^2}{n}$.

$$\sum_{i=1}^{n} Xi^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n} = 91\text{-}73.5 = \mathbf{17.5}$$

Nota: En el Paso 6 se ha calculado la suma de cuadrados de la variable Xi

Paso 7. Eleva al cuadrado cada valor de Yi, y realiza la sumatoria de estos valores;

$\sum_{i=1}^{n} Yi^2$

Datos	
Yi	**Yi^2**
3	$(3)^2 = 9$
4	$(4)^2 = 16$
7	$(7)^2 = 49$
9	$(9)^2 = 81$
8	$(8)^2 = 64$
2	$(2)^2 = 4$
$\sum \mathbf{Yi = 33}$	$\sum_{i=1}^{n} Yi^2 = 223$
$\sum_{i=1}^{n} \frac{(Yi)^2}{n} = \frac{(33)^2}{6} = \frac{1089}{6} = 181.5$	

Paso 8. Eleva al cuadrado la sumatoria de Yi $[(\sum_{i=1}^{n} Yi)^2]$, al resultado es necesario dividirlo entre el tamaño de muestra (n): $\frac{(\sum_{i=1}^{n} Yi)^2}{n}$.

Paso 9. A la sumatoria $\sum_{i=1}^{n} Yi^2$ restarle el resultado de la división $\frac{(\sum_{i=1}^{n} Yi)^2}{n}$.

$$\sum_{i=1}^{n} yi^2 - \frac{(\sum_{i=1}^{n} Yi)^2}{n} = 223\text{-}181.5 = 41.5$$

Nota: En el Paso 9 se ha calculado la suma de cuadrados de la variable Yi

Paso 10. Multiplicar la suma de cuadrados de Xi y de Yi, el resultado de la multiplicación elévalo al exponente un medio ($\frac{1}{2}$) o calcular su raíz cuadrada

$$\left\{\left[\sum_{i=1}^{n} Xi^2 - \frac{(\sum X\,i)^2}{n}\right]\left[\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Yi^2} - \frac{(\sum \boldsymbol{Y\,i})^2}{\boldsymbol{n}}\right]\right\}^{\frac{1}{2}}$$

$$\{[17.5][41.5]\}^{\frac{1}{2}}$$

$$\{726.25\}^{\frac{1}{2}}=26.95$$

Paso 11. Realizar la fórmula para calcular el coeficiente de correlación desacuerdo a la fórmula

$$\boldsymbol{r} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum \mathbf{X\,i})(\sum \mathbf{Y\,i})}{\mathbf{n}}}{\left\{\left[\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathrm{Xi}^2 - \frac{(\sum \mathrm{X\,i})^2}{\mathrm{n}}\right]\left[\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Yi^2} - \frac{(\sum \mathbf{Y\,i})^2}{\mathbf{n}}\right]\right\}^{\frac{1}{2}}}$$

$$\mathrm{r} = \frac{4.5}{26.95}=0.17$$

Literatura citada

Steel GDR, Torrie JH. 1985. Bioestadística: Principios y procedimientos. McGraw-Hill Latinoamericana, ISBN:968-451-495-9: 622 pp.

Dagnino SJ. 2014. Correlación. Revista Chilena de Anestesia. 43: 150-153.

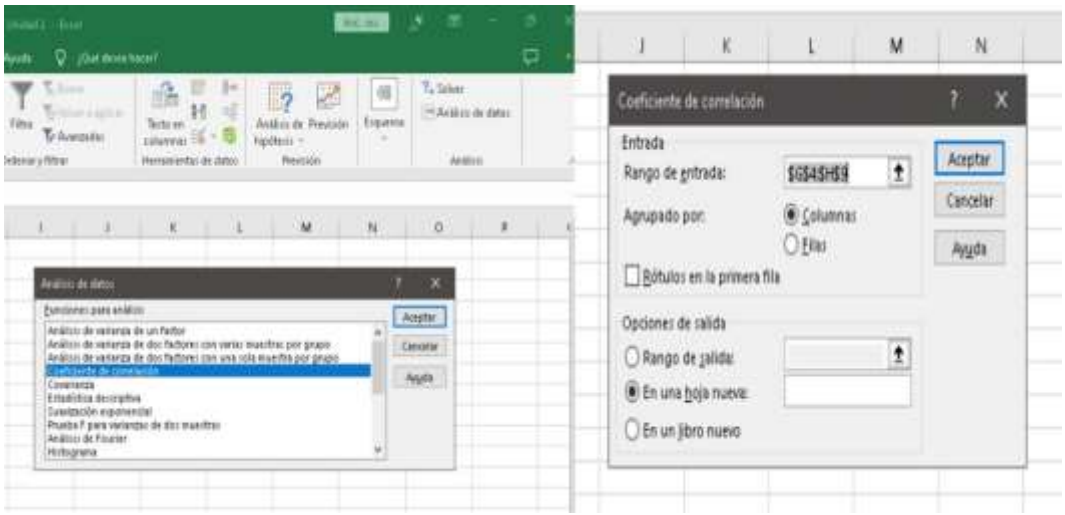
Hernández LJD, Espinosa CJF, Peñaloza TME, Rodríguez JE, Chacón RJG, Toloza SCA, Arenas TMK, Carrillo SSM, Bermúdez PVJ. 2018. Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones. Archivos Venezolanos de Farmacología y Terapéutica 37(5): 587-595.

Lahura E. 2003. El coeficiente de correlación y correlaciones Espúreas. 64 pp. https://core.ac.uk/download/pdf/6445817.pdf

Roy-García I, Rivas-Ruiz R, Pérez-Rodríguez M, Palacios-Cruz L. 2020. Correlación: no toda correlación implica causalidad. Revista Alergia México 66 (3): 354-360.

Tema 12. Correlación en Excel

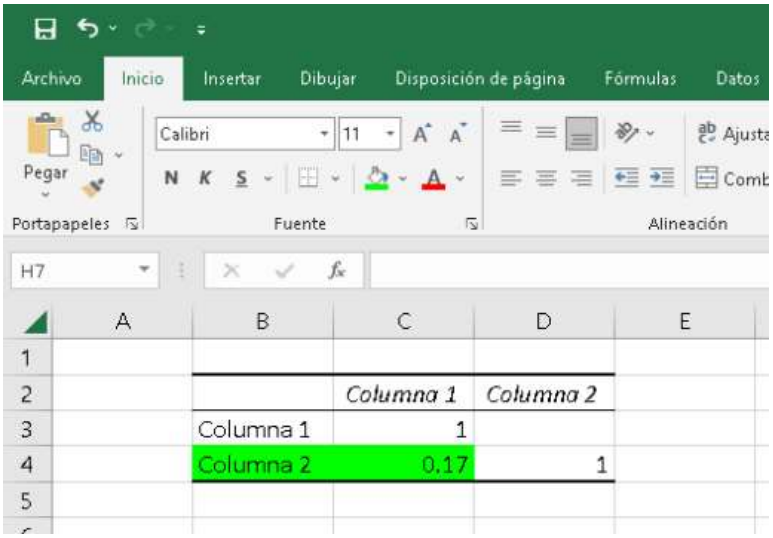
Para realizar una correlación en Excel es necesario **abrir una hoja, ubicar la pestaña de datos** y en la parte final abrir **análisis de datos**, dentro de esta herramienta se **encuentra coeficiente de correlación, la cual se debe seleccionar e introducir los datos de Y y X.**



En la pestaña de **Rango de entrada se introducen los valores de las variables X y Y**, ambos conjuntos de datos deben ser **introducidos sin rótulos**, **se selecciona** la casilla **Agrupado por Columnas**, en las opciones de salida seleccionar **En una hoja nueva**, para que los resultados aparezcan en una hoja nueva.

Resultados del análisis

Los resultados aparecen en una hoja nueva, el coeficiente de correlación es: 0.17.



Conclusión:

En Excel se realiza el análisis con mayor rapidez, el valor del coeficiente de correlación en Excel fue igual que el realizado a mano, el coeficiente de correlación es muy bajo, lo que indica una débil relación en ambas variables.

Tema 13. Correlación en SAS

```
Title"Correlacion";
Data Corr;
Input X Y;
Cards;
1      3
2      4
3      7
4      9
5      8
6      2
;
Proc corr Data=Corr;
Var Y X;
Run;
```

```
                                Correlacion        16:33 Thursday, October 21, 2022   1
                                    Procedimiento CORR
                                2  Variables:    Y        X
                                   Estadísticos simples

                                  Desviación
Variable         N         Media       típica          Suma       Mínimo       Máximo
Y                6       5.50000      2.88097      33.00000      2.00000      9.00000
X                6       3.50000      1.87083      21.00000      1.00000      6.00000


                          Coeficientes de correlación Pearson, N = 6
                               Prob > |r| suponiendo HO: Rho=O

                                                Y            X
                                Y         1.00000      0.16698
                                X         0.16698      1.00000
```

Conclusión:

En SAS el análisis se realiza con mayor rapidez que a mano desarrollando las fórmulas o en Excel, los valores del ejemplo son similares a los obtenidos anteriormente.

Estudio de caso: Regresión lineal y coeficiente de correlación

Instrucciones: Calcula el valor de productos cruzados con las dos fórmulas, el valor de suma de cuadrados (SC) con las tres fórmulas y los coeficientes de regresión lineal (b1 y b0), comprueba la suma de cuadrados de Xi y Yi en el programa SAS, también calcula el coeficiente de correlación y con el valor de coeficiente de correlación concluye la relación de las variables peso de yema y peso de clara.

Valores de Xi (g de yema) y Yi (g de clara)

Xi	10	15	20	25	30
Yi	22	25	27	29	31

Calcula el valor de productos cruzados [PC] con las dos fórmulas:

Fórmula 1:

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{Xi} - \overline{\mathbf{X}})(\mathbf{Yi} - \overline{\mathbf{Y}})$$

Paso 1: Calcular la sumatoria y el promedio de Xi y Yi.

Xi	**Yi**
10	22
15	25
20	27
25	29
30	31
$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n} Yi = 134$
$\overline{X} = \frac{\sum_{i=1}^{n} Xi}{n} = 20$	$\overline{Y} = \frac{\sum_{i=1}^{n} Yi}{n} = 26.8$

Paso 2: Resta a cada valor de Xi y Yi su media respectivamente

Resta $(\text{Xi} - \overline{\text{X}})$

Xi	$(\mathbf{Xi} - \overline{\boldsymbol{X}})$
10	(10 -20) = -10
15	(15 - 20) = -5
20	(20 - 20) = -0
25	(25 - 20) = 5
30	(30 - 20) = 10
$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n} (\text{Xi} - \overline{\text{X}}) = 0$

Resta $(\text{Yi} - \overline{\text{Y}})$

Yi	$(\boldsymbol{Yi} - \overline{\boldsymbol{Y}})$
22	(22 – 26.8) = -4.8
25	(25 – 26.8) = -1.8
27	(27 – 26.8) = 0.2
29	(29 – 26.8) = 2.2
31	(31 – 26.8) = 4.2
$\sum_{i=1}^{n} Yi = 134$	$\sum_{i=1}^{n} (\text{Yi} - \overline{\text{Y}}) = 0$

Paso 3: Multiplica el resultado de la resta de $(Xi - \bar{X})$ por el resultado de la resta de $(Yi - \bar{Y})$.

$(\mathbf{Xi - \bar{X}})$	$(\mathbf{Yi - \bar{Y}})$	$(\mathbf{Xi - \bar{X}})\ (\mathbf{Yi - \bar{Y}})$
(10 -20) = -10	(22 – 26.8) = -4.8	(-10) (-4.8) = 48
(15 - 20) = -5	(25 – 26.8) = -1.8	(-5) (-1.8) = 9
(20 - 20) = -0	(27 – 26.8) = 0.2	(0) (0.2) = 0
(25 - 20) = 5	(29 – 26.8) = 2.2	(5) (2.2) = 11
(30 - 20) = 10	(31 – 26.8) = 4.2	(10) (4.2) = 42
		$\boldsymbol{PC = \sum_{i=1}^{n} (Xi - \bar{X})\ (Yi - \bar{Y}) = 110}$

Fórmula 2:

$$\mathbf{PC = \sum_{i=1}^{n} XiYi - \frac{(\sum_{i=1}^{n} Xi)(\sum_{i=1}^{n} Yi)}{n}}$$

Paso 1: Multiplica los valores de Xi por los valores de Yi

Xi	**Yi**	**XiYi**
10	22	220
15	25	375
20	27	540
25	29	725
30	31	930

$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n} Yi = 134$	$\sum_{i=1}^{n} \mathrm{XiYi} = 2790$
$\overline{\mathrm{X}} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{Xi}}{\mathrm{n}} = 20$	$\overline{\mathrm{Y}} = \frac{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{Yi}}{\mathrm{n}} = 26.8$	

Paso 2: Multiplica las sumatoria $\sum_{i=1}^{n} Xi$ por la sumatoria $\sum_{i=1}^{n} Yi$ y el resultado de la multiplicación dividirlo entre el tamaño de muestra; $\frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}}$

$$\frac{(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Xi})(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Yi})}{\mathbf{n}} = \frac{(\mathbf{100})(\mathbf{134})}{\mathbf{5}} = \frac{\mathbf{13400}}{\mathbf{5}} = \mathbf{2680}$$

Paso 3: Calcula el valor de productos cruzado con la fórmula:

$$\mathbf{PC} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{XiYi} - \frac{(\sum_{i=1}^{n} \boldsymbol{Xi})(\sum_{i=1}^{n} \boldsymbol{Yi})}{\mathbf{n}} = \mathbf{2790} - \mathbf{2680} = \mathbf{110}$$

Calcula la Suma de Cuadrados Corregidos de la variable Xi con las tres fórmulas siguientes:

Fórmula 1:

$$\sum_{i=1}^{n} (\mathrm{Xi} - \overline{\mathrm{X}})^2$$

Paso 1: Realiza la sumatoria de Xi y calcula su media

Paso 2: Resta a cada valor de Xi su media; $(\mathbf{Xi} - \overline{\boldsymbol{X}})$

Paso 3: Eleva al cuadrado el resultado de la resta suma los valores; $\sum_{i=1}^{n}(\mathrm{Xi} - \bar{X})^2$

Xi	$(\mathbf{Xi} - \bar{\boldsymbol{X}})$	$(\mathbf{Xi} - \bar{\boldsymbol{X}})^2$
10	(10 -20) = -10	100
15	(15 - 20) = -5	25
20	(20 - 20) = -0	0
25	(25 - 20) = 5	25
30	(30 - 20) = 10	100
$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n}(\mathrm{Xi} - \bar{\mathrm{X}}) = 0$	$\boldsymbol{\sum_{i=1}^{n}(\mathbf{Xi} - \bar{\mathbf{X}})^2 = 250}$

Formula 2:

$$\sum_{i=1}^{n} \mathrm{Xi}^2 - \frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n}$$

Paso 1. Cada valor de Xi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^2$.

Paso 2. Calcular la sumatoria de Xi $\left[\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Xi}\right]$ elévala al cuadrado y divídela entre n [n= tamaño de la muestra]; $\frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n}$

Paso 3. Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^2$ **menos** $\frac{\left(\sum_{i=1}^{n} Xi\right)^2}{n}$

Xi	$\boldsymbol{Xi^2}$
10	100
15	225
20	400

25	625
30	900
$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 2250$
$\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Xi^2}\text{-}\frac{(\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{Xi})^2}{\mathbf{n}} = 2250 - \frac{(100)^2}{5} = 2250 - 2000 = \mathbf{250}$	

Fórmula 3:

$$\left[\sum_{i=1}^{n} \boldsymbol{Xi^2} - \boldsymbol{n\bar{X}^2}\right]$$

Paso 1. Cada valor de Xi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Xi}^2$

Paso 2: Multiplica el tamaño de muestra (n) por la media elevada al cuadrado $\bar{\mathrm{X}}^2$

Paso 3: Realiza la resta de $\sum_{i=1}^{n} \mathrm{Xi}^2$ **menos** $n\bar{X}^2$

Xi	$\boldsymbol{Xi^2}$
10	100
15	225
20	400
25	625
30	900
$\sum_{i=1}^{n} Xi = 100$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 2250$
$\sum_{i=1}^{n} \mathrm{Xi}^{\mathbf{2}} - n\bar{X}^2 = 2250 - 5(20^2) = 2250\text{-}\mathbf{2000} = \mathbf{250}$	

Análisis de la variable Xi para encontrar la Suma de Cuadrados corregidos

```
Title"SCxx";
Data SCxx;
Input X;
Cards;
10
15
20
25
30
;
Proc Univariate;
Var X;
Run;
```

Resultados en SAS

SCxx 14:42 Wednesday, February 24, 2023 5

Procedimiento UNIVARIATE

Variable: X

Momentos

N	5	Pesos de la suma	5
Media	20	Observaciones de la suma	100
Desviación típica	7.90569415	Varianza	62.5
Suma de cuadrados no corregidos	**2250**	**Suma de cuadrados corregidos**	**250**
Coeficiente de variación	39.5284708	Media de error estándar	3.53553391

Conclusión:

La suma de cuadrados corregidos en SAS coincide con los resultados encontrados con las fórmulas anteriores: 250

Calcula b1 con las tres fórmulas anteriores

Fórmula1:

$$\frac{\sum_{i=1}^{n}(Xi-\bar{X})(Yi-\bar{Y})}{\sum_{i=1}^{n}(Xi-\bar{X})^2}=\frac{110}{250}=0.44$$

Fórmula 2:

$$\frac{\sum_{i=1}^{n}XiYi-\frac{(\sum_{i=1}^{n}Xi)(\sum_{i=1}^{n}Yi)}{n}}{\sum_{i=1}^{n}Xi^2-\frac{(\sum_{i=1}^{n}Xi)^2}{n}}=\frac{110}{250}=0.44$$

Fórmula 3:

$$\frac{\sum_{i=1}^{n}XiYi-\frac{(\sum_{i=1}^{n}Xi)(\sum_{i=1}^{n}Yi)}{n}}{\sum_{i=1}^{n}Xi^2-n\bar{X}^2}=\frac{110}{250}=0.44$$

Calcula b0:

Fórmula:

$$b0=\bar{Y}-b1\bar{X}$$

$$b0=26.8-20(0.44)=26.8-8.8=18$$

Análisis de regresión para las variables peso de yema y peso de clara

```
Title"Regresion";
Data Reg;
Input X Y;
Cards;
10    22
15    25
20    27
25    29
30    31
;
Proc REG;
Model Y=X;
Run;
```

Parámetros estimados

Variable	DF	Parameter Estimate	Standard Error	Valor t	Pr > |t|
Término i	1	**18.00000**	0.48990	36.74	<.0001
Xdependie	1	**0.44000**	0.02309	19.05	0.0003

Conclusión:

Los coeficientes de regresión bo y b1, son similares a los calculados con las fórmulas anteriores [b0=18 y b1=0.44].

Calcula la Suma de Cuadrados corregidos de la variable Yi con las tres fórmulas siguientes:

Fórmula 1:

$$\sum_{i=1}^{n}(\mathbf{Yi} - \bar{\boldsymbol{Y}})^2$$

Paso 1: Realiza la sumatoria de Yi y calcula su media

Paso 2: Resta a cada valor de Yi su media; $(\mathbf{Yi} - \bar{\boldsymbol{Y}})$

Paso 3: Eleva al cuadrado el resultado de la resta suma los valores; $\sum_{i=1}^{n}(\mathrm{Yi} - \bar{Y})^2$

Yi	$(\mathbf{Yi} - \bar{\boldsymbol{Y}})$	$(\mathbf{Yi} - \bar{\boldsymbol{Y}})^2$
22	(22 – 26.8) = -4.8	23.04
25	(25 – 26.8) = -1.8	3.24
27	(27 – 26.8) = 0.2	0.04
29	(29 – 26.8) = 2.2	4.84
31	(31 – 26.8) = 4.2	17.64
$\sum_{i=1}^{n} Yi = 134$	$\sum_{i=1}^{n}(Yi - \bar{Y}) = 0$	$\sum_{i=1}^{n}(\boldsymbol{Yi} - \bar{\boldsymbol{Y}})^2 = \mathbf{48.8}$

Formula 2:

$$\sum_{i=1}^{n} \mathrm{Yi}^2 - \frac{\left(\sum_{i=1}^{n} Yi\right)^2}{n}$$

Paso 1. Cada valor de Yi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Yi}^2$.

Paso 2. Calcular la sumatoria de Xi $\left[\sum_{i=1}^{n} Yi\right]$ elévala al cuadrado y divídela entre n [n= tamaño de la muestra]; $\frac{\left(\sum_{i=1}^{n} Yi\right)^2}{n}$

Paso 3. Realiza la resta de $\sum_{i=1}^{n} \mathrm{Yi}^2$ **menos** $\frac{\left(\sum_{i=1}^{n} Yi\right)^2}{n}$

Yi	$\boldsymbol{Yi^2}$
22	484
25	625
27	729
29	841
31	961
$\sum_{i=1}^{n} Xi = 134$	$\sum_{i=1}^{n} \mathrm{Xi}^2 = 3640$
$\boldsymbol{\sum_{i=1}^{n} \mathrm{Yi}^2 - \frac{\left(\sum_{i=1}^{n} \mathrm{Yi}\right)^2}{n}} = 3640 - \frac{(134)^2}{5} = 3640 - 3591.2 = \mathbf{48.8}$	

Fórmula 3:

$$\left[\sum_{i=1}^{n} \boldsymbol{Yi^2} - \boldsymbol{n\bar{Y}^2}\right]$$

Paso 1. Cada valor de Yi elévalo al cuadrado y realiza la sumatoria de estos; $\sum_{i=1}^{n} \mathrm{Yi}^2$

Paso 2: Multiplica el tamaño de muestra (n) por la media elevada al cuadrado $\bar{Y}^2$

Paso 3: Realiza la resta de $\sum_{i=1}^{n} \mathrm{Yi}^2$ **menos** $n\bar{\mathrm{Y}}^2$

Yi	Yi^2
22	484
25	625
27	729
29	841
31	961
$\sum_{i=1}^{n} Yi = 134$	$\sum_{i=1}^{n} \mathrm{Yi}^2 = 3640$
$\sum_{i=1}^{n} \mathrm{Yi}^2 - n\overline{\mathrm{Y}}^2 = 3640 - 5(26.8^2) = 3640$-**3591.2= 48.8**	

Análisis de la variable Yi para encontrar la Suma de Cuadrados corregidos

```
Title"SCyy";
Data SCyy;
Input Y;
Cards;
22
25
27
29
31
;
Proc Univariate;
Var y;
Run;
```

```
                                SCyy          14:43 Wednesday, February 24, 2023   9

                              Procedimiento UNIVARIATE
                                   Variable:  Y
                                     Momentos
N                                          5    Pesos de la suma                          5
Media                                   26.8    Observaciones de la suma                134
Desviación típica                3.49284984    Varianza                               12.2
Suma de cuadrados no corregidos          3640    Suma de cuadrados corregidos           48.8
Coeficiente de variación        13.0330218    Media de error estándar          1.56204994
```

Calcular el coeficiente de correlación (r) de las variables anteriores [peso de yema y peso de clara]

Fórmula 1:

$$r=\frac{\sum_{i=1}^{n}(Xi-\overline{x})(Yi-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(Xi-\overline{X})^2\sum_{i=1}^{n}(Yi-\overline{Y})^2}}$$

$$r=\frac{110}{\sqrt{(250)(48.8)}}=\frac{110}{\sqrt{12200}}=\frac{110}{110.45}=0.995$$

Fórmula 2:

$$r=\frac{\sum_{i=1}^{n}XiYi-\frac{(\sum_{i=1}^{n}Xi)(\sum_{i=1}^{n}Yi)}{n}}{\sqrt{\left[\sum_{i=1}^{n}Xi^2-\frac{(\sum_{i=1}^{n}Xi)^2}{n}\right]\left[\sum_{i=1}^{n}Yi^2-\frac{(\sum_{i=1}^{n}Yi)^2}{n}\right]}}=0.995$$

Conclusión:

Las variables están altamente correlacionadas, el coeficiente de correlación fue de 0.995

Análisis de correlación de las variables peso de yema y peso de clara

```
Title"Correlacion";
Data Corr;
Input X Y;
Cards;
10 22
15 25
20 27
25 29
30 31
;
Proc corr Data=Corr;
Var Y X;
Run;
```

Coeficientes de correlación Pearson, N = 5
Prob > |r| suponiendo H0: Rho=0

	Y	X
Y	1.00000	0.99589 0.0003
X	0.99589 0.0003	1.00000

Conclusión:

El coeficiente de correlación en SAS es similar al calculado con las fórmulas anteriores.

Las variables están altamente correlacionadas, el coeficiente de correlación fue de 0.995

Tema 14. Diseño Completamente al Azar [DCA]

Este diseño es el más sencillo, eficiente y se origina por la asignación aleatoria de los tratamientos a un conjunto de unidades experimentales previamente determinado (Badii *et al.*, 2007). Este diseño tiene amplia aplicación cuando las unidades experimentales son homogéneas. La homogeneidad de las unidades experimentales puede lograrse ejerciendo un control local apropiado (seleccionando: sujetos, animales o plantas de una misma edad, raza, variedad o especie). Pero debe tenerse presente que todo material biológico, por homogéneo que sea, presenta una cierta fluctuación cuyos factores no se conocen y son, por lo tanto, incontrolables.

El nombre diseño completamente al azar deriva del hecho que existe completamente una aleatorización, la cual valida la prueba F de Fisher-Snedecor. También se le conoce como Diseño de una vía o un sólo criterio de clasificación en virtud de que las respuestas se hallan clasificadas únicamente por los tratamientos.

Modelo

El modelo lineal es una expresión algebraica que condensa todos los factores presentes en la investigación:

$$\boldsymbol{Y_{ij}} = \mu + \boldsymbol{T_i} + \boldsymbol{E_{ij}}$$

Donde:

$Y_{ij} =$ variable de interes

μ= es la media general del experimento

T_i= es el efecto del tratamiento

E_{ij}=es el error aleatorio asociado a la respuesta Y_{ij}

Cuadro 1. Análisis de varianza de un diseño completamente al azar.

Fuente de variación	GL	SC	CM	F-calculada	F-tablas
Tratamiento (trat)	**t-1**	$\frac{\sum_{i=1}^{n} Y^2_{i.}}{r} - \frac{\sum_{i=1}^{n} Y_{..}^{\ 2}}{tr}$	$\frac{SC\ trat}{GL\ trat}$	$\frac{CM\ trat}{CM\ error}$	
Error	**t (r-1)**	**SC total - SC trat**	$\frac{SC\ error}{GL\ error}$		
Total (tot)	**tr-1**	$\sum_{i=1}^{n} Y_{ij}^{\ 2} - \frac{\sum_{i=1}^{n} Y_{..}^{\ 2}}{tr}$			

***GL: grados de libertad, **SC: suma de cuadrados, ***CM: cuadrados medios**

Conclusión:

Es necesario comparar la F-calculada con la F-tablas. Si la F-calculada es mayor que la F-tablas se rechaza la hipótesis nula* (H0) *y se acepta* (H1), a*l menos un tratamiento es distinto.

Operaciones para realizar la suma de cuadrados del tratamiento, error y total

Suma de cuadrados del tratamiento

Pasos para calcular la *SC trat* (*suma de cuadrados del tratamiento*)

1.- Suma las repeticiones de cada tratamiento

2.- Cada una de las sumas elévalas al cuadrado

3.- Suma todos los valores elevados al cuadrado anterior mente

4.- La sumatoria anterior divídela entre las repeticiones (r)

5.- Al resultado de la división se le resta el factor de corrección (FC)

Fórmula uno:

$$\frac{\sum_{i=1}^{n} Y^2{}_{i.}}{r} - \frac{\sum_{i=1}^{n} Y^2{}_{..}}{tr}$$

Fórmula dos:

$$\frac{\sum_{i=1}^{n} Y^2{}_{i.}}{r} - \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}\right)^2}{tr}$$

Pasos para calcular el Factor de Corrección

1.- Suma las repeticiones de cada tratamiento

2. Suma cada una de las sumatorias anteriores y el resultado elevarlo al cuadrado

3.- Divide el resultado anterior entre la multiplicación tratamiento por repetición (tr)

$$\frac{\sum_{i=1}^{n} Y^2{}_{..}}{tr}$$

Otra fórmula para calcular el factor de corrección

$$\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}\right)^2}{tr}$$

Esta fórmula indica suma todas las repeticiones de cada tratamiento y ese valor elevarlo al cuadrado y el resultado dividelo entre la multiplicación del número de tratamientos por las repeticiones.

Suma de cuadrados del total

1.- Eleva cada valor de las repeticiones al cuadrado

2.- Suma cada valor elevado al cuadrado (Y^2ij)

3.- Al resultado de la sumatoria restar el FC

$$\sum_{i=1}^{n} Y_{ij}{}^2 - FC$$

Suma de cuadrados del error

La suma de cuadrados del error se calcula por diferencia:

$$SC\ error = SC\ total - SC\ tratamiento$$

Estudio de caso: diseño completamente al azar

El dueño de un corral de un engorda de bovinos, desea probar tres desparasitantes para controlar garrapatas, por lo que decidió dividir sus becerros en tres tratamientos [cada desparasitante es un tratamiento], cada tratamiento con cinco repeticiones [cada corral es una repetición, cada corral con 10 toros]. ¿Existe diferencia en la efectividad de los desparasitantes?

Tratamiento 1	Tratamiento 2	Tratamiento 3
10	8	2
13	7	3
10	9	4
11	7	3
17	8	2

La pregunta que hace el dueño, ¿existe alguna diferencia entre los desparasitantes?, al dueño le interesa saber la eficacia de los desparasitantes para poder aplicar el mejor en todo su rebaño.

Calcular el valor de la SC trat.

1.- Suma las repeticiones de cada tratamiento:

Tratamiento 1	Tratamiento 2	Tratamiento 3
10	8	2
13	7	3
10	9	4
11	7	3
17	8	2
$\sum_{i=1}^{n} 61$	$\sum_{i=1}^{n} 39$	$\sum_{i=1}^{n} 14$

2.- Las sumatorias de las repeticiones elevarlas al cuadrado

$\sum_{i=1}^{n}(61)^2=3721$	$\sum_{i=1}^{n}(39)^2=1521$	$\sum_{i=1}^{n}(14)^2=196$

3.- Suma todos los valores elevados al cuadrado anterior mente

$$\sum_{i=1}^{n} Y^2{}_{i.} = 3721 + 1521 + 196 = 5438$$

4.- La sumatoria anterior divídela entre las repeticiones (r)

$$\frac{\sum_{i=1}^{n} Y^2{}_{i.}}{r} = \frac{5438}{5} = 1087.6$$

5.- Al resultado de la división réstale el factor de corrección (FC)

$$\sum_{i=1}^{n} Y^2{}_{i.} - \frac{\sum_{i=1}^{n} Y^2{}_{..}}{tr}$$

Calcula el valor del Factor de corrección

1.- Suma las repeticiones de cada tratamiento

Tratamiento 1	Tratamiento 2	Tratamiento 3
10	8	2
13	7	3
10	9	4
11	7	3
17	8	2
$\sum_{i=1}^{n} 61$	$\sum_{i=1}^{n} 39$	$\sum_{i=1}^{n} 14$

2. Suma cada una de las sumatorias anteriores y ese valor elevarlo al cuadrado

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}\right)^2 = \sum_{i=1}^{n} Y_{..}^{\,2} = (61+39+14)^2 = (114)^2 = 12996$$

3.- Divide el resultado anterior entre la multiplicación tratamiento por repetición (tr)

$$\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} Y_{ij}\right)^2}{tr} = \frac{\sum_{i=1}^{n} Y_{..}^{\,2}}{tr} = \frac{12996}{15} = 866.4$$

Suma de cuadrados del tratamiento:

$$\sum_{i=1}^{n} Y^2{}_{i.} - \frac{\sum_{i=1}^{n} Y_{..}^{\,2}}{tr} = 1087.6\text{-}866.4 = \mathbf{221.2}$$

Calcula la Suma de cuadrados del total

1.- Eleva cada valor de las repeticiones al cuadrado

Trat-1	Y_{ij}^2	Trat-2	Y_{ij}^2	Trat- 3	Y_{ij}^2
10	**100**	8	**64**	2	**4**
13	**169**	7	**49**	3	**9**
10	**100**	9	**81**	4	**16**
11	**121**	7	**49**	3	**9**
17	**289**	8	**64**	2	**4**
	$\sum_{i=1}^{n} Y_{ij}^2 = 779$		$\sum_{i=1}^{n} Y_{ij}^2 = 307$		$\sum_{i=1}^{n} Y_{ij}^2 = 42$

2.- Suma cada valor elevado al cuadrado (Y_{ij}^2)

$$\sum_{i=1}^{n} Y_{ij}^2 = 779+307+42 = 1128$$

3.- Al resultado de la sumatoria restar el FC

$$\sum_{i=1}^{n} Y_{ij}^2 - \frac{\sum_{i=1}^{n} Y_{..}^2}{tr} = 1128 - 866.4 = 261.6$$

Calcula la Suma de cuadrados del error

La suma de cuadrados del error se calcula por diferencia:

$$\boldsymbol{SC\ error = SC\ total - SC\ tratamiento = 261.6 - 221.2 = 40.4}$$

Una vez calculadas las sumas de cuadrados del: tratamiento, error y total, el paso siguiente es calcular los cuadrados medios (CM) y el valor de la F calculada (Cuadro 2).

Cuadro 2. ***Análisis de varianza del diseño completamente al azar, estudio de caso: tres desparasitantes, cinco repeticiones (corral), diez toros por repetición.***

Fuente de variación	GL	SC	CM	F-calculada
Tratamiento	2	221.2	$\frac{221.2}{2} = 110.6$	$\frac{110.6}{3.4} = 32.5$
Error	12	40.4	$\frac{40.4}{12} = 3.4$	
Total	14	261.6	$\frac{SCtotal}{Gl\ Total} = S^2$	

Es necesario encontrar el valor de F tablas y comparar con el valor de F calculada y con concluir (Cuadro 3).

F Grados de libertad del tratamiento

Grados de libertad del error, α

En la tabla de distribución de F con un α=0.05 se busca el valor donde se intercepten los grados de libertad del tratamiento (primera fila) y los grados de libertad del error (primera columna).

Ejemplo: para $n_1 = 5$, $n_2 = 10$ y $\alpha = 0.05$, $F_{5,10;0.05} = 3.326$, significa que $P(F_{5,10} > 3.326) = 0.05$.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	12	15	16	18	20	24
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	246.5	247.3	248.0	249.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.43	19.44	19.45	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.692	8.675	8.660	8.639
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.844	5.821	5.803	5.774
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.604	4.579	4.558	4.527
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.922	3.896	3.874	3.841
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.494	3.467	3.445	3.410
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.202	3.173	3.150	3.115
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.989	2.960	2.936	2.900
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.828	2.798	2.774	2.737
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.701	2.671	2.646	2.609
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.599	2.568	2.544	2.505
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.515	2.484	2.459	2.420
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.445	2.413	2.388	2.349
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.385	2.353	2.328	2.288
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.333	2.302	2.276	2.235

El valor de F de tablas con un α=0.05 es de 3.885.

Cuadro 3. *Comparación de F tablas con la F calculada*

F calculada	*Signo*	*F tablas*
32.5	***>* *(mayor que)***	***3.885***

Conclusión

La F calculada es mayor que la F de tablas, por lo que se rechaza Ho (hipótesis nula) y se acepta H1 (hipótesis alternativa) con un α=0.05. Al menos uno de los desparasitantes tiene un efecto distinto a los otros dos.

Literatura citada

Badii, MH, Castillo J, Rodríguez M, Wong A, Villalpando P.2007. Diseños experimentales e investigación científica. Innovaciones de Negocios 4(2): 283–330.

Segura CJC. 2000. Notas de diseños Experimentales. Universidad Autónoma de Yucatán. Facultad de Medicina Veterinaria y Zootecnia. 54 pp.

Análisis de varianza de todos datos con PROC UNIVARIATE

Al dividir la suma de cuadrados del total entre los grados de libertad del total el valor resultante es la varianza de todos los datos.

$$\frac{SC\ total}{Gl\ total} = \frac{261.6}{14} = 18.68$$

Comprobando este valor en el programa SAS, se aplicó el procedimiento PROC UNIVARIATE.

```
Input Trat Rep Gpta;
Cards;
1 1 10
1 2 13
1 3 10
1 4 11
1 5 17
2 1 8
2 2 7
2 3 9
2 4 7
2 5 8
3 1 2
3 2 3
3 3 4
3 4 3
3 5 2
;
Proc univariate;
Var Gpta;
Run;
```

RESULTADOS

Medidas estadísticas básicas

Localización		Variabilidad	
Media	7.600000	Desviación típica	4.32270
Mediana	8.000000	**Varianza**	**18.68571**
Moda	2.000000	Rango	15.00000

Conclusión:

El análisis de varianza de todos los datos fue similar al resultado encontrado al dividir la suma de cuadrados del total entre los grados de libertad del total.

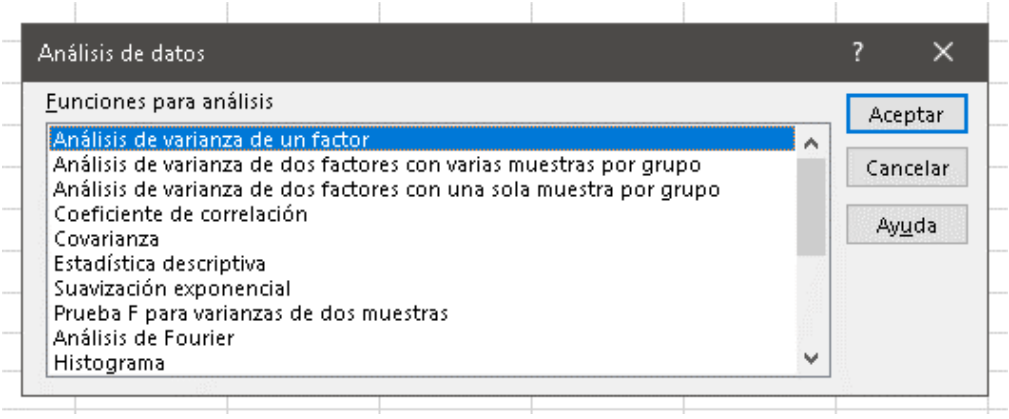
Tema 15. Diseño completamente al azar en Excel

En Excel los datos son escritos de forma vertical, las repeticiones de cada tratamiento en una columna. En la barra de herramientas dar click en datos, entrar en análisis de datos.

Cuadro 1. Tres tratamientos (Trat; desparasitantes), cinco repeticiones (cada repetición es un corral), cada repetición con 10 toros, los valores representan el número de garrapatas después de aplicar el desparasitante.

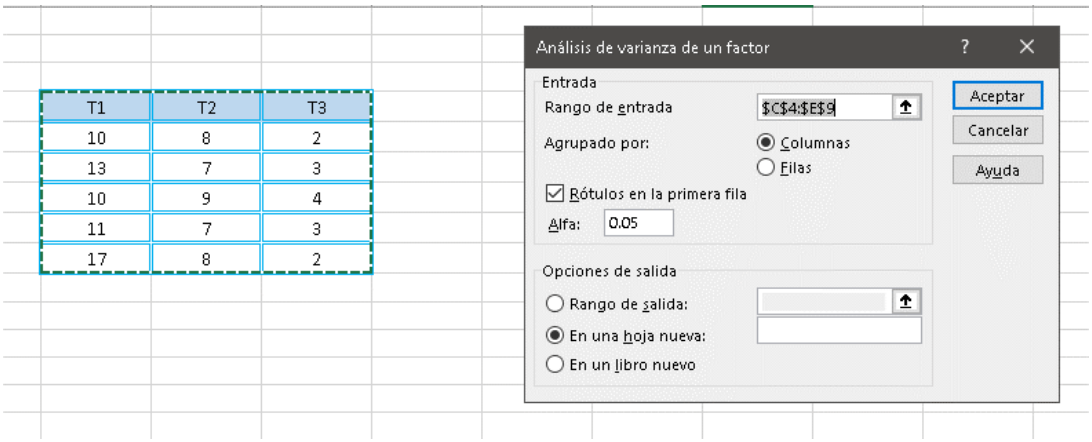
Trat 1	Trat 2	Trat 3
10	8	2
13	7	3
10	9	4
11	7	3
17	8	2

Al abrir una hoja en Excel, en la barra de herramientas Datos, entrar en Análisis de datos, en el catálogo de funciones para análisis seleccionar: Análisis de varianza de un factor.



Al seleccionar análisis de varianza de un factor, en la pestaña Rango de entrada seleccionas todos los datos con los rótulos de tratamiento (T), seleccionar agrupador por: Columna,

Rótulos en la primera fila, con un alfa: 0.05 y finalmente seleccionar Opciones de salida en una hoja nueva.



Los resultados del análisis aparecen en una hoja nueva, los resultados observados en los grados de libertad, la suma de cuadrados, la F calculada son similares a los encontrados al realizar las operaciones con las fórmulas, el valor de la F tablas es similar al encontrado en la tabla de Fisher con un alfa igual a 0.05.

Análisis de varianza de un factor

RESUMEN

Grupos	Cuenta	Suma	Promedio	Varianza
T1	5	61	12.2	8.7
T2	5	39	7.8	0.7
T3	5	14	2.8	0.7

ANÁLISIS DE VARIANZA

Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Entre grupos	221.2	2	110.6	32.85140515	1.35663E-05	3.88529383
Dentro de los grupos	40.4	12	3.366666667			
Total	261.6	14				

Conclusión: el análisis de varianza de un diseño completamente al azar se puede realizar en SAS y en Excel.

Tema 16. Prueba de Tukey en SAS

Un ganadero desea conocer la efectividad de tres desparasitantes contra garrapatas, distribuyo sus toros en tres tratamientos (tratamientos=desparasitantes), cinco repeticiones (cada repetición es un corral), cada repetición con 10 toros, los valores representan el número de garrapatas después de aplicar el desparasitante. A los datos se aplicará una prueba de TUKEY para conocer cuál es el desparasitante que reduce la infestación de garrapatas.

```
Data DCA;
Input Trat Rep Gpta;
Cards;
1 1 10
1 2 13
1 3 10
1 4 11
1 5 17
2 1 8
2 2 7
2 3 9
2 4 7
2 5 8
3 1 2
3 2 3
3 3 4
3 4 3
3 5 2
;
Proc GLM;
Class Trat Rep;
Model Gpta= Trat;
Means Trat/tukey;
Run;
```

Sunday, April 4, 2023

RESULTADOS

Procedimiento GLM

Información del nivel de clase

Clase	Niveles	Valores
Trat	3	1 2 3
Rep	5	1 2 3 4 5

Número de observaciones 15

Los resultados coinciden con el diseño completamente al azar; tres tratamientos, cinco repeticiones, quince observaciones en total.

Análisis de varianza en SAS

Sistema SAS 18:34 Sunday, April 4, 2023
Procedimiento GLM

Variable dependiente: Gpta

Fuente	DF	Suma de cuadrados	Cuadrado de la media	F-Valor	Pr > F
Modelo	2	221.2000000	110.6000000	**32.85**	<.0001
Error	12	40.4000000	3.3666667		
Total correcto	14	261.6000000			

Los resultados del análisis de varianza en SAS coinciden con los observados calculados a mano y en Excel.

Conclusión: Al menos un desparasitante presenta mayor eficacia en el control de garrapatas.

Comparación de medias por TUKEY

Sistema SAS 18:34 Sunday, April 4, 2023
Procedimiento GLM

Prueba del rango estudentizado de Tukey (HSD) para Gpta

NOTA: Este test controla el índice de error experimentwise de tipo I, pero normalmente tiene un

índice de error de tipo II más elevado que REGWQ.

Alfa	0.05
Error de grados de libertad	12
Error de cuadrado medio	3.366667
Valor crítico del rango estudentizado	3.77278
Diferencia significativa mínima	3.0958

Medias con la misma letra no son significativamente diferentes.

Tukey Agrupamiento	Media	N	Trat
A	12.200	5	1
B	7.800	5	2
C	**2.800**	**5**	**3**

Conclusión:

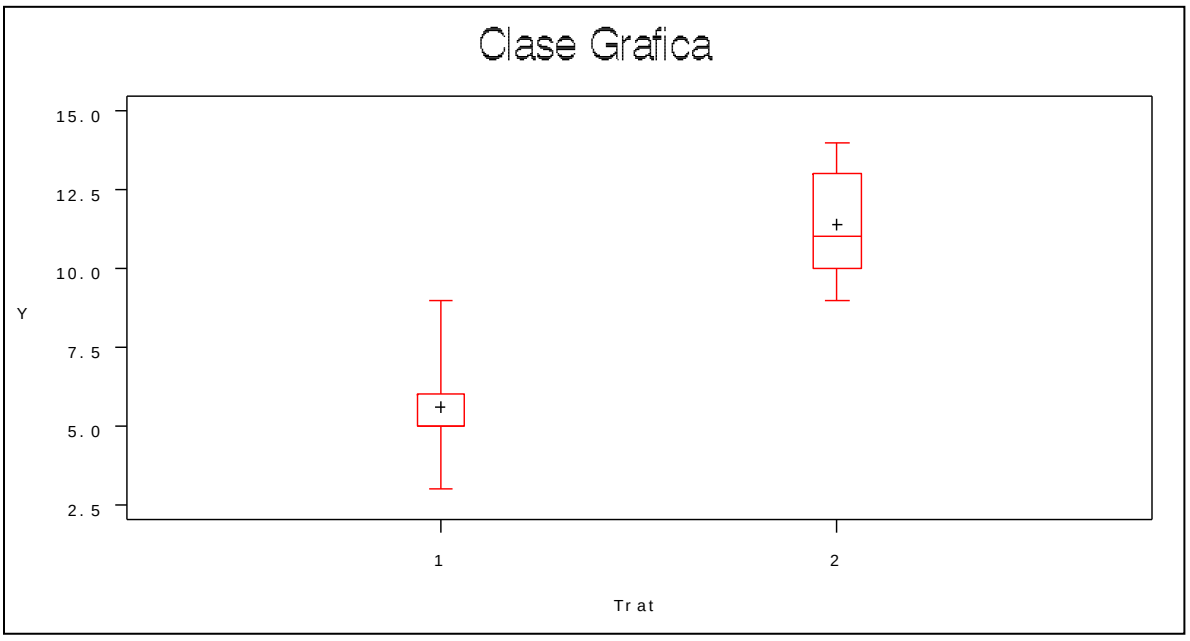
Al aplicar la comparación de medias por TUKEY se observó que el tratamiento tres presento la mayor efectividad, el número de garrapatas que se observaron después de su aplicación fue la menor

Tema17. Proc Boxplot (Gráfica de cajas en SAS)

El procedimiento Proc Boxplot genera una gráfica de cajas. En el siguiente ejemplo, se muestra el procedimiento para graficar el promedio de la variable Y en dos tratamientos.

```
Title"Clase Grafica";
Data Grafica;
Input Trat Y;
Cards;
1 3
1 6
1 5
1 5
1 9
2 9
2 10
2 11
2 13
2 14
;
/*Proc Print;
Run;*/
Proc Boxplot data=Grafica;
Plot Y*TRAT;
Run;
```

Resultados de la salida



Tema 18: PROC FREQ (análisis de frecuencia en SAS)

En este ejemplo se desea saber el porcentaje de pollos nacidos con cresta rosa y con cresta simple. El uso de PROC FREQ clasifica en porcentaje la presencia de cresta simple y cresta rosa de la siguiente base de datos.

```
Data Cresta;
Input Pollo Peso Cresta$;
Cards;
1     35    S
2     35    R
3     35    S
4     35    S
5     35    R
6     35    S
7     35    R
8     35    S
9     35    R
10    55    R
11    30    S
12    40    S
13    40    S
14    40    S
15    40    S
16    35    S
17    35    S
18    35    R
19    35    R
20    35    S
21    35    S
22    35    R
23    35    S
24    35    R
25    35    R
26    60    S
27    60    R
28    40    R
29    40    R
30    40    R
31    40    S
32    35    S
33    35    S
34    35    R
35    35    R
36    35    R
37    35    S
38    35    R
39    35    S
40    35    R
41    35    S
42    35    R
```

```
43      35      S
44      35      S
45      35      S
46      35      R
47      35      S
48      35      S
49      35      S
50      35      S
51      35      R
52      35      S
53      35      S
54      35      S
55      35      R
56      35      R
57      35      S
58      35      S
59      35      S
60      35      R
61      35      R
62      35      R
63      35      S
64      35      S
65      35      S
66      35      S
67      35      S
68      35      R
69      35      S
70      35      R
71      35      R
72      35      S
73      35      R
;
Proc Freq Data=cresta;
Table cresta / chisq;
Run;
```

RESULTADOS

```
Sistema SAS          14:22 Tuesday, March 30, 2021   3

                        Procedimiento FREQ
                                                 Frequencia     Porcentaje
                Cresta     Frecuencia     Porcentaje     acumulada     acumulado
                R                  31          42.47            31          42.47
                S                  42          57.53            73         100.00

                        Test chi-cuadrado
                      para proporciones de igualdad
                      Chi-cuadrado          1.6575
                      DF                         1
                      Pr > ChiSq            0.1979
                        Tamaño de la muestra = 73
```

Tema 19. Prueba de t-Sudent de dos muestras suponiendo varianzas iguales

La prueba t de Student fue diseñada por un matemático llamado W. Gosset, se la utiliza cuando la prueba de hipótesis implica una comparación entre dos medias muéstrales. La expresión distribución t designa una familia de distribuciones teóricas que sirven a la prueba de hipótesis, cuando las muestras son pequeñas (N≤30; Sánchez, 2015).

En la prueba t, los grados de libertad se representan como n-1 (número de casos menos uno). Dado un valor determinado de α y los grados de libertad, se compara el resultado del valor t obtenido con el valor tabulado. Si el valor t calculado es igual o mayor que el valor de la t de tablas puede rechazarse la H0 al correspondiente nivel de significación indicado.

Cálculo del estadístico t calculada (t_c)

$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p{}^2 \left(\frac{1}{n1} + \frac{1}{n2}\right)}}$$

Donde:

$\overline{X}_1$ y $\overline{X}_2$ = Son las medias de las muestras observadas

n1= Tamaño de la muestra uno

n2= Tamaño de la muestra dos

S_p^2= Varianza común (promedio ponderado de las dos varianzas de las muestras)

Fórmula dos para encontrar el valor de t_c:

$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{EED}$$

Donde:

$\overline{X}_1$ y $\overline{X}_2$ = Son las medias de las muestras observadas

EED= Error estándar de la diferencia de medias

Para encontrar el valor de EED se han desarrollado dos fórmulas:

Fórmula 1 para encontrar el valor de EED

$$EED = \sqrt{{S_P}^2 \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}$$

Fórmula 2 para encontrar el valor de EED:

$$EED = \sqrt{{S_P}^2 \left[\frac{n_1 + n_2}{(n_1)(n_2)}\right]}$$

Fórmula de la varianza común:

$${S_P}^2 = \frac{\left[(n_1 - 1){S_1}^2\right] + \left[(n_2 - 1){S_2}^2\right]}{n_1 + n_2 - 2}$$

Donde:

S_p^2= Varianza común

n1= Tamaño de la muestra uno

n2= Tamaño de la muestra dos

S_1^2= Varianza de la muestra uno (X_1)

S_2^2= Varianza de la muestra dos (X_2)

Como calcular la t tablas (t_t): los grados de libertad: $gl=n1+n2-2$, la elección del nivel de significancia (α) depende de la magnitud del error que se desea asumir. También depende del nivel de confianza con el que quieras trabajar (α-1= nivel de confianza).

Conclusión se determina de la comparación:

$t_{calculada}$	Mayor que $>$	t_{tablas}

Si T-Calculada es mayor que T-tablas se rechaza H0 [hipótesis nula]

Pasos para encontrar el valor de t-calculada

1.- Encontrar el valor de la media de X_1 y de X_2.

2. Calcular el valor de las varianzas de la muestra uno y de la muestra dos [S_1^2 y S_1^2].

3. Calcular los grados de libertad de la muestra uno y de la muestra dos (fórmula de los grados de libertad (*gl*); [*gl* = n-1].

Literatura citada

Lugo-Armenta JG, Pino-Fan LR. 2021. Niveles de Razonamiento Inferencial para el Estadístico t-Student. Bolema, Rio Claro 35 (71): 1776-1802.

Ortiz JE, Moreno EC. 2011. ¿Se necesita la prueba t de Student para dos muestras independientes asumiendo varianzas iguales? Comunicaciones en Estadística 4 (2): 139-157.

Pastor-Barriuso R. 2012. Bioestadística. Madrid: Centro Nacional de Epidemiología, Instituto de Salud Carlos III. 251 pp.

Sánchez TRA 2015. t-Student. Usos y abusos. Revista Mexicana de Cardiología. 26 (1): 59-61.

Estudio de caso: "Efectividad de dos medicamentos contra coriza aviar"

Instrucciones: dos medicamentos [A y B] fueron aplicados a muestras de pollos [n=20 y n=18], con diagnóstico de una misma enfermedad [coriza aviar, ocasionada por: *Avibacterium paragallinarum*]. Se midió la presencia de glóbulos blancos que están relacionados con la efectividad del medicamento. El avicultor desea conocer si existe diferencia en la efectividad de los medicamentos.

Obs	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Medicamentos																				
A	6	5	6	7	5	7	6	4	3	6	6	5	7	8	6	5	8	4	6	6
B	7	6	7	9	5	8	7	6	7	9	8	7	8	7	6	8	7	6		

Obs: observaciones

Muestra X_1

Medicamentos																			
A	6	5	6	7	5	7	6	4	3	6	6	5	7	8	6	5	8	4	6	6

1. Sumatoria

$$\sum_{i=1}^{n} X_1 = 116$$

2. Promedio

$$\overline{X}_1 = \frac{116}{20} = 5.8$$

3. Varianza

$$S_1{}^2 = \frac{(Xi - \overline{X})^2}{n - 1} = 1.642$$

Muestra X_2

Medicamentos																			
B	7	6	7	9	5	8	7	6	7	9	8	7	8	7	6	8	7	6		

1. Sumatoria

$$\sum_{i=1}^{n} X_2 = 128$$

2. Promedio

$$\overline{X}_2 = \frac{116}{20} = 7.1$$

3. Varianza

$$S_2{}^2 = \frac{(Xi - \overline{X})^2}{n - 1} = 1.163$$

4.-Calcular la varianza común:

$$S_P{}^2 = \frac{\left[(n_1 - 1){S_1}^2\right] + \left[(n_2 - 1){S_2}^2\right]}{n_1 + n_2 - 2}$$

$$S_P{}^2 = \frac{[20 - 1]1.642 + [18 - 1]1.163}{20 + 18 - 2}$$

$$S_P{}^2 = \frac{31.20 + 19.78}{36} = \frac{50.98}{36} = 1.41$$

5.- Encontrar el valor de EED:

$$EED = \sqrt{1.41\left[\frac{20 + 18}{(20)(18)}\right]} = \sqrt{\frac{1.41}{1}\left[\frac{38}{360}\right]}$$

$$EED = \sqrt{\frac{1.41}{1}\left[\frac{38}{360}\right]} = \sqrt{\frac{53.80}{360}} = \sqrt{0.149} = 0.38$$

6.- Calcular el valor de t-calculada (t_c)

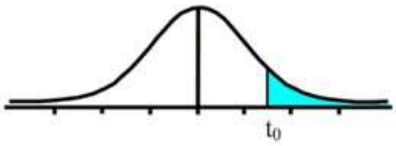
$$t_c = \frac{\overline{X}_1 - \overline{X}_2}{EED}$$

$$t_c = \frac{5.8 - 7.1}{0.38} = \frac{-1.3}{0.38} = -3.39$$

Encontrar el valor de t de tablas (t_t)

- Grados de libertad [gl]= n1+n2 - 1 = 20 + 18 – 2=38-2=36
- Probabilidad ($\left(1-\frac{\alpha}{2}\right)$) = 0.975

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079

El valor de t de tablas (t_t) = $\mathbf{2.0281}$

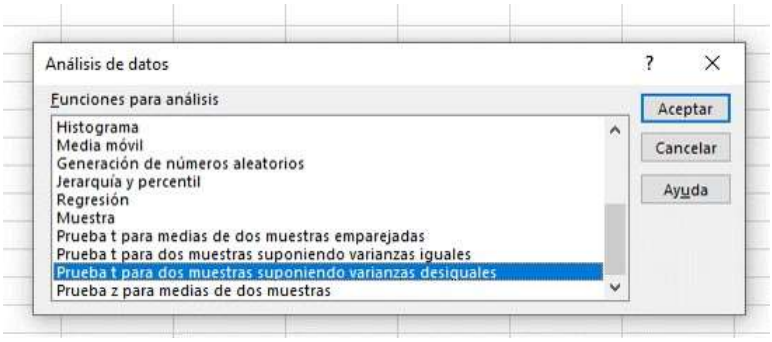
Conclusión se determina de la comparación:

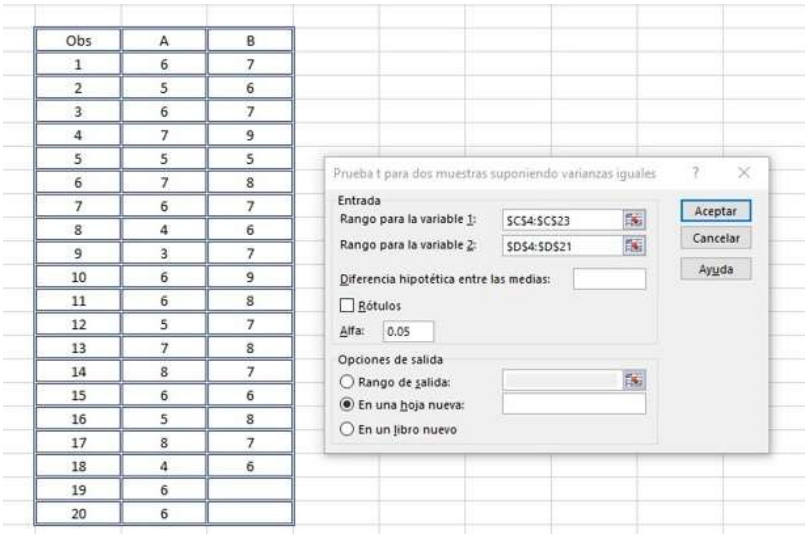
$t_{calculada}$	Mayor que (>), menor que (<)	t_{tablas}
-*3.39*	Mayor que	2.0281

T-calculada es mayor que t-tablas se rechaza H0 [hipótesis nula]. Se acepta la hipótesis alternativa, los medicamentos presentan diferente efectividad contra coriza aviar.

Tema 20. Prueba de t de Sudent en Excel

La prueba de T-student se puede realizar en Excel, es necesario abrir una hoja, seleccionar la herramienta datos y seleccionar análisis de datos. En análisis de datos seleccionar prueba de t para dos muestras suponiendo varianzas iguales.





Los resultados de la prueba de t aparecen en una hoja nueva:

Prueba t para dos muestras suponiendo varianzas iguales		
	Variable 1	*Variable 2*
Media	5.8	7.1111111
Varianza	1.64210526	1.1633987
Observaciones	20	18
Varianza agrupada	**1.41604938**	
Diferencia hipotética de las medias	0	
Grados de libertad	36	
Estadístico t	**-3.3912496**	
P(T<=t) una cola	0.0008512	
Valor crítico de t (una cola)	**1.68829771**	
P(T<=t) dos colas	0.0017024	
Valor crítico de t (dos colas)	**2.028094**	

Conclusion:

Los resultados son similares a los encontrados en el ejemplo resuelto con las fórmulas, el valor del estadistico t_c (-3.39) es mayor que el estadistico t_t (2.028), esta comparación indica que se rechaza la hipótesis nula, los medicamentos tienen diferente efectividad contra coriza aviar.