Prueba de hipótesis: Chi-cuadrada

UNIVERSIDAD LATINA DE COSTA RICA
LAUREATE INTERNATIONAL UNIVERSITIES®

- *Estadística paramétrica*: se aplica básicamente a variables continuas. Consiste en el conjunto de técnicas estadísticas de estimación de parámetros, intervalos de confianza y pruebas de hipótesis. Se asume que la población cumple los principios de una distribución normal o se aproxima.

- *Estadística no paramétrica*: consiste en la comparación de distribuciones, ya que no es posible determinar la distribución original ni la distribución de los estadísticos, por lo que no se tiene parámetros a estimar.

Chi-cuadrada

- La prueba de Chi-cuadrada es utilizada para analizar la frecuencia de dos variables con categorías múltiples, permite determinar si existe independencia entre las mismas o lo contrario. Proporciona una prueba para contrastar frecuencias observadas contra esperadas.

  Ejemplo:

  ¿El tipo de refresco preferido por un grupo de personas es independiente de su categoría etaria?

  ¿La preferencia política está asociada al género de una persona?

  ¿Se relaciona el tipo de película que ve una persona con la edad?

  ¿Está relacionada la profesión que decide estudiar un estudiante con el contexto social donde vive?

Prueba de Chi-cuadrada.

- Paso 1. Plantear la prueba de hipótesis.

  H_0: Las variables son INDEPENDIENTES.

  H_1: Las variables NO son INDEPENDIENTES.

- Paso 2. Definir el nivel de significancia.
- Paso 3. Desarrollar el estadístico de prueba.
- Paso 4. Determinar el valor crítico.
- Paso 5. Tomar la decisión (aceptar o rechazar la H_0).
- Paso 6. Elaborar la conclusión.

Caso de estudio.

- El dueño de una pequeña empresa que fabrica jabones para el lavado de ropa, ha lanzado al mercado un nuevo jabón de baño en tres presentaciones diferentes.
- El jabón se vende en tres tiendas departamentales dentro de la ciudad y el dueño está interesado en saber, si el número de jabones que se venden (de cada presentación) podría estar relacionado con la tienda departamental en las que se ofrecen. Trabaje con un 90% de confianza.

Los datos de las ventas del último mes aparecen en la siguiente tabla:

Caso de estudio.

Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra
Todo Mart	1220	460	845
Tiendas Ahorrará	1204	503	890
Comercial del Abarrote	1280	456	820

Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
Todo Mart	1220	460	845	2525
Tiendas Ahorrará	1204	503	890	2597
Comercial del Abarrote	1280	456	820	2556
Totales	3704	1419	2555	7678

Caso de estudio.

- Paso 1. Plantear la prueba de hipótesis.

  H_0: *Las variables son INDEPENDIENTES.*

  H_1: *Las variables NO son INDEPENDIENTES.*

- Paso 2. Definir el nivel de significancia.

  $\alpha = 0{,}1$

- Paso 3 y 4. Desarrollar el estadístico de prueba y determinar el valor crítico.

UNIVERSIDAD LATINA DE COSTA RICA
LAUREATE INTERNATIONAL UNIVERSITIES®

Caso de estudio.

SUMA | =B$5*($E2/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo M	=B$5*($E2/$E$5)		840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					

SUMA | =B$5*($E3/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas	=B$5*($E3/$E$5)		864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					

Caso de estudio.

SUMA | =C$5*($E2/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart		=C$5*($E2/$E$5)		
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					

SUMA | =C$5*($E4/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote		=C$5*($E4/$E$5)		
11					

Caso de estudio.

SUMA | =D$5*($E2/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,	=D$5*($E2/$E$5)	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					

SUMA | =D$5*($E4/$E$5)

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,	=D$5*($E4/$E$5)	
11					

Caso de estudio.

SUMA | =(B14-C14)^2/C14

	A	B	C	D	E
1	Tienda departamental	Jabón líquido	Jabón en polvo	Jabón en barra	Totales
2	Todo Mart	1220	460	845	2525
3	Tiendas Ahorrará	1204	503	890	2597
4	Comercial del Abarrote	1280	456	820	2556
5	Totales	3704	1419	2555	7678
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					
12					
13		Observados	Esperados	Chi- cuadrada	
14	Todo Mart	1220	1218	=(B14-C14)^2/C14	
15	Todo Mart	460	466,6547	0,0949	
16	Todo Mart	845	840,2416	0,0269	
17	Tiendas Ahorrará	1204	1252,8377	1,9038	
18	Tiendas Ahorrará	503	479,9613	1,1059	
19	Tiendas Ahorrará	890	864,2010	0,7702	
20	Comercial del Abarrote	1280	1233,0586	1,7870	
21	Comercial del Abarrote	456	472,3840	0,5683	
22	Comercial del Abarrote	820	850,5574	1,0978	
23		Valor Chi cuadrada de prueba:		7,3577	
24		Valor Chi cuadrada crítica:		7,7794	

SUMA | =SUMA(D14:D22)

	A	B	C	D	E
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					
12					
13		Observados	Esperados	Chi- cuadrada	
14	Todo Mart	1220	1218,1037	0,0030	
15	Todo Mart	460	466,6547	0,0949	
16	Todo Mart	845	840,2416	0,0269	
17	Tiendas Ahorrará	1204	1252,8377	1,9038	
18	Tiendas Ahorrará	503	479,9613	1,1059	
19	Tiendas Ahorrará	890	864,2010	0,7702	
20	Comercial del Abarrote	1280	1233,0586	1,7870	
21	Comercial del Abarrote	456	472,3840	0,5683	
22	Comercial del Abarrote	820	850,5574	1,0978	
23		Valor Chi cuadrada de		=SUMA(D14:D22)	
24		Valor Chi cuadrada crítica:		SUMA(**número1**; [número2]; ...)	
25		Grados de libertada:		4	GL= (# (

SUMA | =PRUEBA.CHI.INV(0,1;4)

	A	B	C	D	E
6					
7	Tienda departamental	Jabón Líquido	Jabón en polvo	Jabón en barra	
8	Todo Mart	1218,1037	466,6547	840,2416	
9	Tiendas Ahorrará	1252,8377	479,9613	864,2010	
10	Comercial del Abarrote	1233,0586	472,3840	850,5574	
11					
12					
13		Observados	Esperados	Chi- cuadrada	
14	Todo Mart	1220	1218,1037	0,0030	
15	Todo Mart	460	466,6547	0,0949	
16	Todo Mart	845	840,2416	0,0269	
17	Tiendas Ahorrará	1204	1252,8377	1,9038	
18	Tiendas Ahorrará	503	479,9613	1,1059	
19	Tiendas Ahorrará	890	864,2010	0,7702	
20	Comercial del Abarrote	1280	1233,0586	1,7870	
21	Comercial del Abarrote	456	472,3840	0,5683	
22	Comercial del Abarrote	820	850,5574	1,0978	
23			Valor Chi cuadrada de prueba:	7,3577	
24			Valor Chi cuadrada (	=PRUEBA.CHI.INV(0,1;4)	
25			Grados de libertada:	**PRUEBA.CHI.INV**(probabilidad; grad	

22	Comercial del Abarrote	820	850,5574	1,0978	
23		Valor Chi cuadrada de prueba:		7,3577	
24		Valor Chi cuadrada crítica:		7,7794	
25		Grados de libertada:		4	GL= (# de columnas -1) * (# filas -1)
26					

Caso de estudio.

- Paso 5 y 6. Tomar la decisión y elaborar una conclusión.

  Como el estadístico de prueba es menor al estadístico crítico, se puede concluir que no hay evidencia para rechazar la hipótesis de independencia, por lo que se puede decir que las ventas de jabón no dependen de la tienda departamental en la que son vendidos.

Caso de estudio.

- Un productor de fármacos ha fabricado la presentación de un medicamento para la gripe, y ha lanzado al mercado el producto a través de tres farmacias diferentes.
- El dueño está interesado en saber, si el número de medicamentos que se venden (de cada presentación) podría estar relacionado con las farmacias en las que se ofrecen. Trabaje con un 90% de confianza.

Los datos de las ventas del último mes aparecen en la siguiente tabla:

Farmacia	Presentación A	Presentación B	Presentación C
La Económica	1732	1512	1008
El Glovo	1755	1776	2086
Michel	2870	2155	1339

Análisis de correlación lineal y regresión lineal simple.

UNIVERSIDAD LATINA DE COSTA RICA
LAUREATE INTERNATIONAL UNIVERSITIES®

Asociación estadística entre variables

- Según el nivel de medición de las variables en cuestión así se mide el grado de la asociación estadística entre ellas.

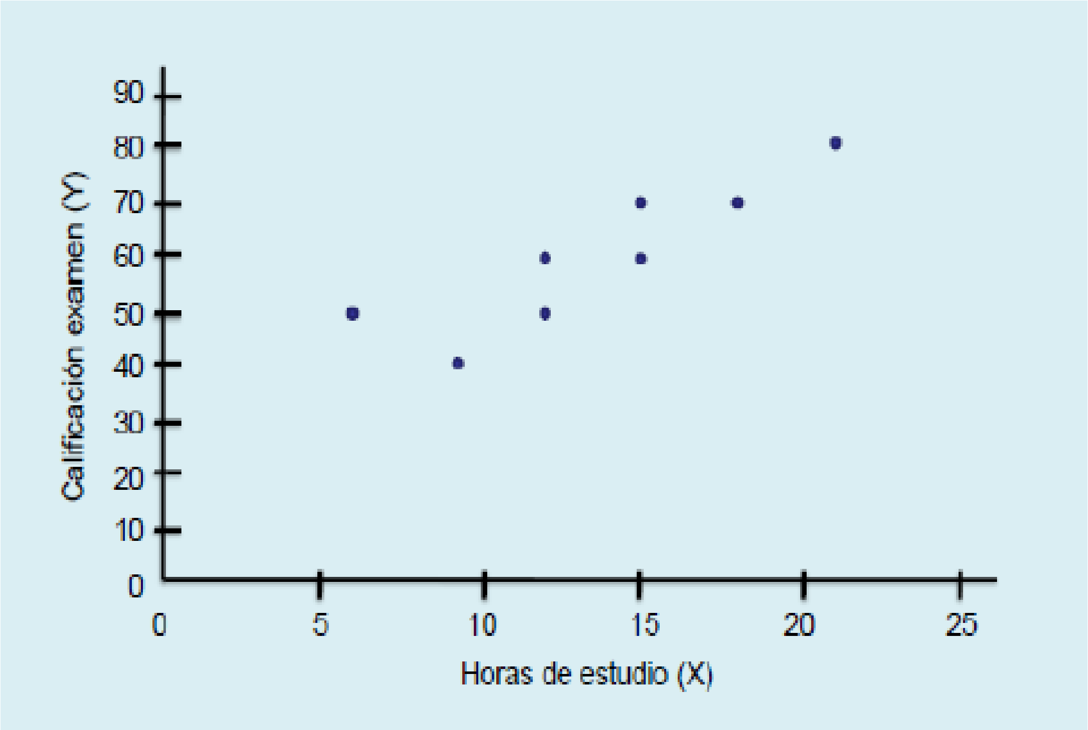
- Cuando se tienen dos variables cuantitativas resulta necesario conocer cómo es la relación establecida en cuanto a dos aspectos principales:

  La naturaleza o dirección de la asociación entre las variables.
  La fuerza o intensidad de la asociación entre las variables.

Asociación estadística entre variables

Correlación: Relación recíproca entre dos variables.

Diagrama de dispersión: gráfico que utiliza las coordenadas cartesianas para mostrar la relación entre dos variables cuantitativas.



Ejemplo de aplicación.

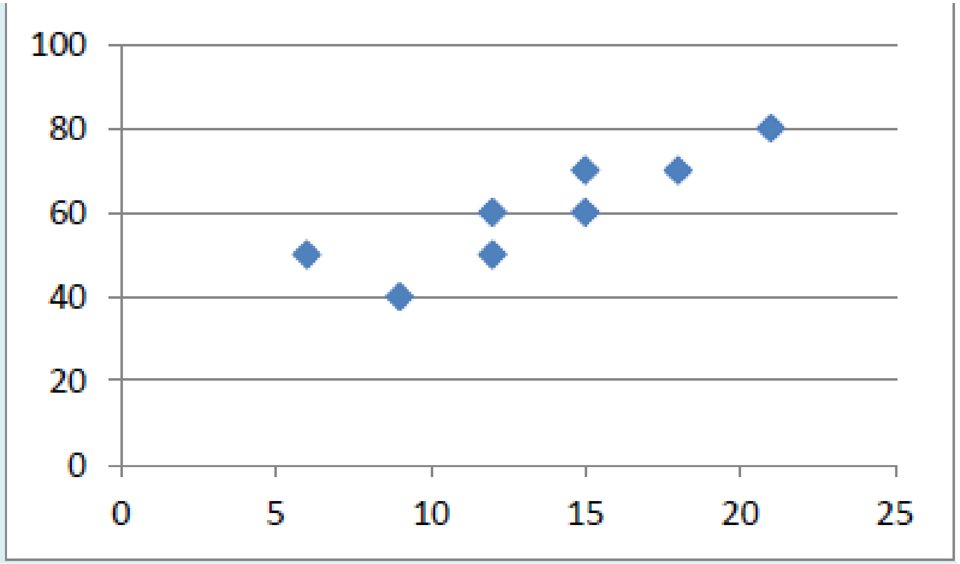
Un investigador desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba. ¿Cuáles son las variables de este problema y cuáles son los datos que el investigador podría buscar?

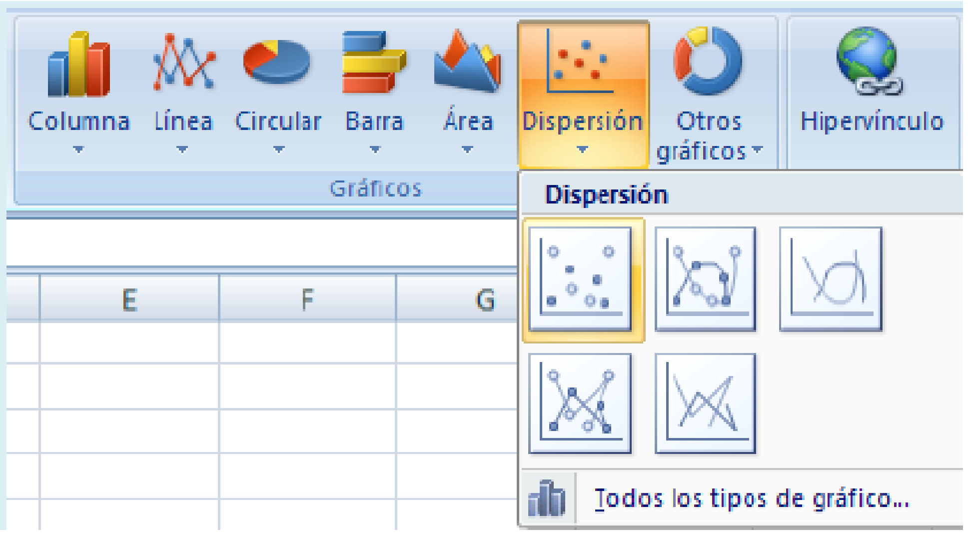
Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

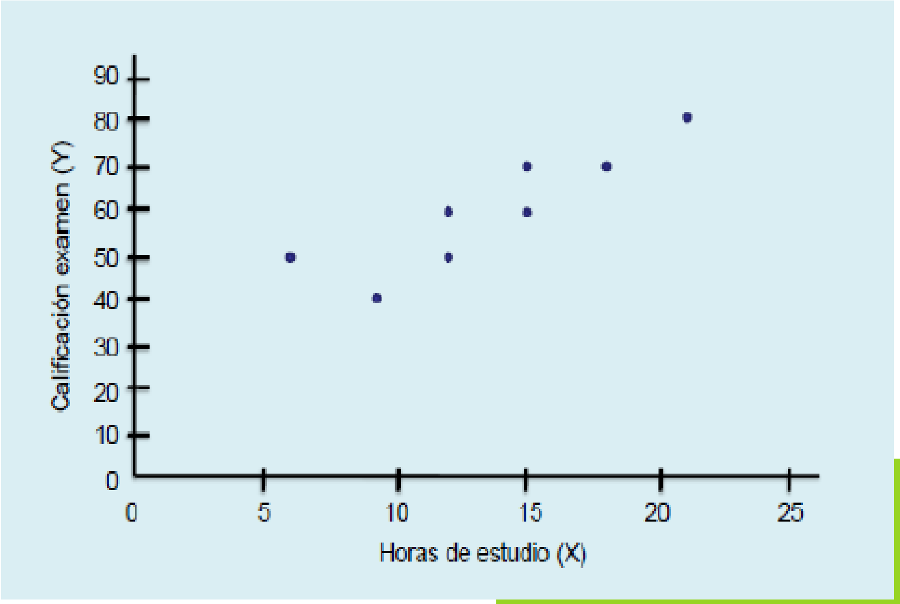
UNIVERSIDAD LATINA DE COSTA RICA
LAUREATE INTERNATIONAL UNIVERSITIES®

Ejemplo de aplicación.

	A	B
1	X	Y
2	21	80
3	15	60
4	15	70
5	9	40
6	12	60
7	18	70
8	6	50
9	12	50







Coeficiente de correlación lineal de Pearson.

- Existen diferentes maneras de calcular el coeficiente de correlación lineal de Pearson. Talvez la más sencilla es empelar la fórmula de cálculo siguiente:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

N = número de pares de valores X y Y observados

SXY = sumatoria de todos los valores de X

SY = sumatoria de todos los valores de Y

SX2 = es la sumatoria de los cuadrados de cada uno de los valores de Y

Ejemplo de aplicación.

Un investigador desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba. ¿Cuáles son las variables de este problema y cuáles son los datos que el investigador podría buscar?

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ	108 ↑ ΣX	480 ↑ ΣY	6870 ↑ ΣXY	1620 ↑ ΣX^2	30000 ↑ ΣY^2

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

$N = 8,\ \Sigma X = 108,\ \Sigma Y = 480,\ \Sigma XY = 6870,\ \Sigma X^2 = 1620,\ \Sigma Y^2 = 30000:$

$$r = \frac{8 \cdot 6870 - 108 \cdot 480}{\sqrt{(8 \cdot 1620 - (108)^2)(8 \cdot 30000 - (480)^2)}}$$

$$r = 0{,}885$$

Regresión lineal simple.

- Cuando se quiere establecer una relación lineal entre variables con un nivel de correlación aceptable, entonces se puede emplear la técnica de la regresión lineal simple para determinar la ecuación de una recta que permita pronosticar el comportamiento de y en términos de x.
- **Regresión lineal simple:** es un método matemático que modela la relación lineal entre una variable dependiente Y y una variable independiente X.

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum Y}{N} - b\frac{\sum X}{N}$$

Ejemplo de aplicación.

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ	108 ↑ ΣX	480 ↑ ΣY	6870 ↑ ΣXY	1620 ↑ ΣX^2	30000 ↑ ΣY^2

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{8 \cdot 6870 - 108 \cdot 480}{8 \cdot 1620 - (108)^2}$$

$$b = 2{,}41$$

$$a = \frac{\sum Y}{N} - b\frac{\sum X}{N}$$

$$a = \frac{480}{8} - 2{,}41\frac{108}{8}$$

$$a = 27{,}5$$

$$y = 27{,}5 + 2{,}41x$$

O bien, Nota= 27,5 + 2,41 * Número de horas de preparación.

Bondad de ajuste: El coeficiente de determinación.

- Cuando se ha construido un modelo de regresión, resultado importante contar con una medida de que tan bien se ajusta la recta estimada a los datos observados. Esa medida es el coeficiente de determinación R^2. El coeficiente de determinación se calcula como el cuadrado del coeficiente de correlación de Pearson.

$$R^2 = r^2$$

- Entre más cercano a uno, indica un mejor ajuste de la recta. Se puede interpretar como la proporción de la variabilidad explicada por el modelo.

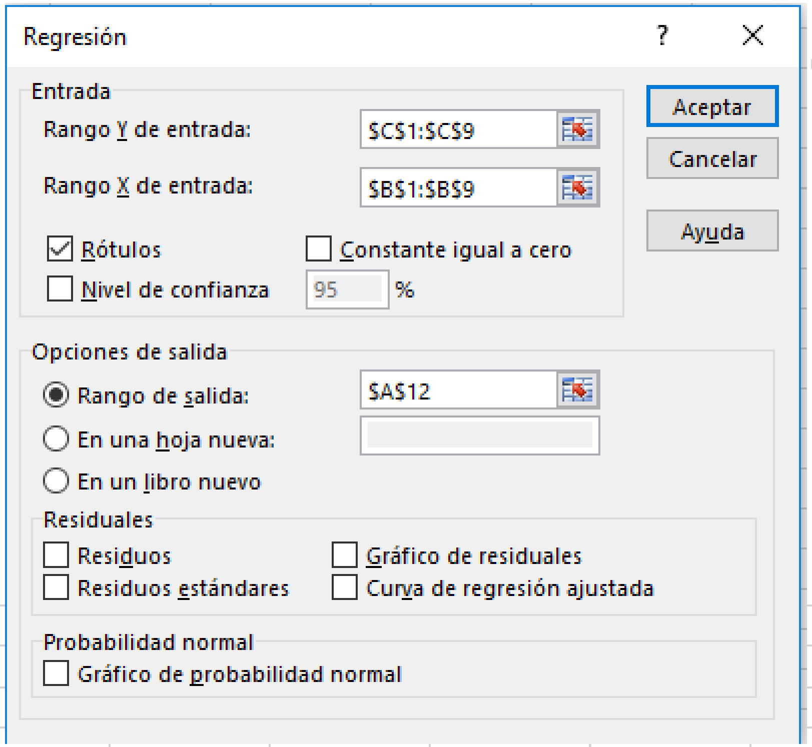
Ejemplo de aplicación.

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ	108 ↑ ΣX	480 ↑ ΣY	6870 ↑ ΣXY	1620 ↑ ΣX^2	30000 ↑ ΣY^2

$$R^2 = r^2 = (0,885)^2 = 0,7832$$

Este resultado quiere decir que el modelo de regresión planteado explica el 78,32% de la variabilidad de *y*, o sea, que la relación lineal entre la nota en el examen de estadística y el número de horas de preparación explica el 78,32% de la variabilidad de las notas. Este valor de *R*2 indicaría que es un modelo bastante bueno, pues posee un poder explicativo alto.

Estudiante	Horas de estudio (x)	Nota obtenida (y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50



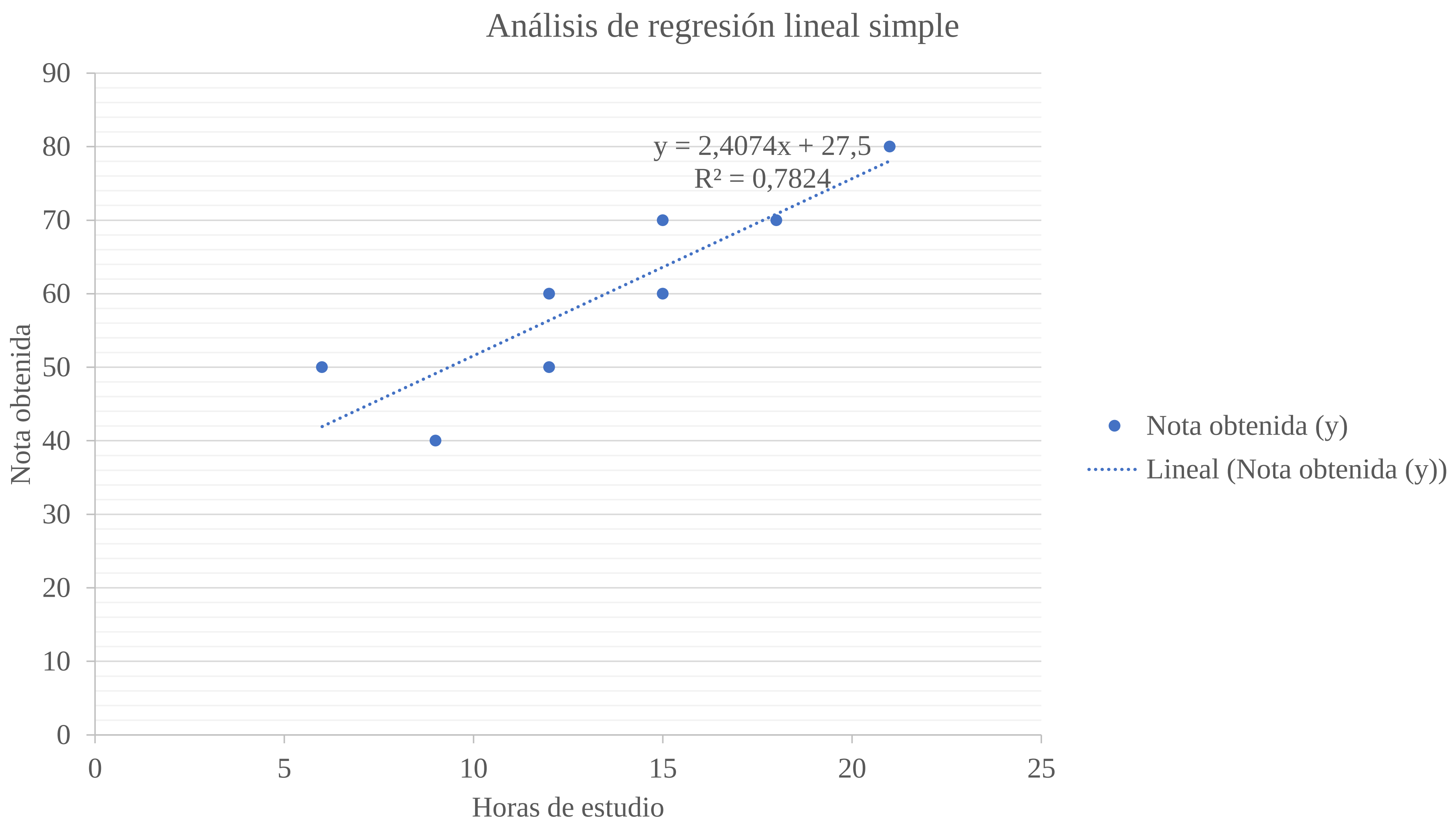
Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,884537963
Coeficiente de determinación R^2	0,782407407
R^2 ajustado	0,746141975
Error típico	6,596856715
Observaciones	8

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	lio de los cua	F	alor crítico de F
Regresión	1	938,8888889	938,888889	21,5744681	0,00352265
Residuos	6	261,1111111	43,5185185		
Total	7	1200			

	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	nferior 95,0%	uperior 95,0%
Intercepción	27,5	7,375510026	3,72855571	0,00975329	9,45277711	45,5472229	9,45277711	45,5472229
Horas de estudio (x)	2,407407407	0,518298013	4,64483241	0,00352265	1,13917786	3,67563696	1,13917786	3,67563696


Análisis de regresión lineal simple
y = 2,4074x + 27,5
R² = 0,7824
Nota obtenida
90
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
Horas de estudio
Nota obtenida (y)
Lineal (Nota obtenida (y))

Bibliografía

- Rodríguez Franco, Jesús, Pierdant Rodríguez, Alberto Isaac. Estadística para administración. (Primera Edición). México: Grupo Editorial Patria. (2014).
- Leandro Oviedo Gabriel. Estadística y Probabilidad con aplicaciones. (Primera Edición). Costa Rica: Publitex Grupo Editorial S.A. (2014).