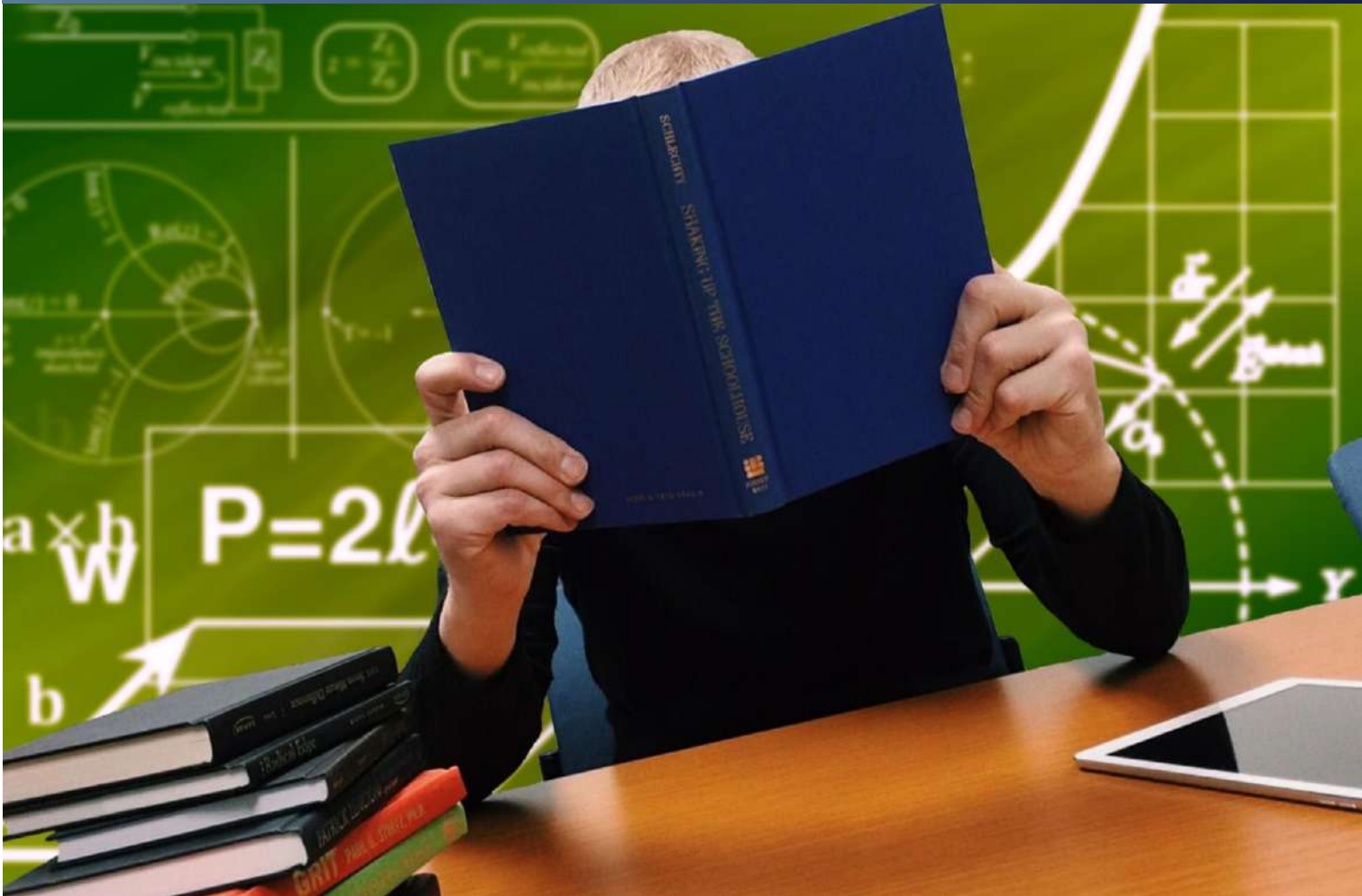


# *Ejercicios y Talleres*



puedes enviarlos a  
[klasesdematematicasymas@gmail.com](mailto:klasesdematematicasymas@gmail.com)

## EJERCICIOS DE REGRESIONES Y ANALISIS DE COVARIANZA

Analizar la información recopilada por medio de los diferentes métodos para poder tomar decisiones, realizando proyecciones futuras por medio de series cronológicas.

### PREGUNTAS ABIERTAS:

1. En la tabla dada a continuación se informa el número de consultas por día, atendidas a los estudiantes de la USTA, en educación a distancia.

Número de consultas	Frecuencia
0 -5	5
6- 10	9
11 - 15	8
16 - 20	4
21- 25	1
26 - 30	1

Calcule la varianza, la desviación estándar y encuentre un intervalo de confianza de 90%.

Número de consultas	Frecuencia $f_i$	Marca de clase $\hat{y}_i$	$\hat{y}_i \cdot f_i$	$(\hat{y}_i - \bar{x})^2 \cdot f_i$
0 -5	5	2,5	12,5	337,372449
6- 10	9	7,5	67,5	92,9846939
11 - 15	8	12,5	100	25,5102041
16 - 20	4	17,5	70	184,183673
21- 25	1	22,5	22,5	138,903061
26 - 30	1	27,5	27,5	281,760204
Total	<b>28</b>		<b>300</b>	<b>1060,71429</b>

$$\text{Media} = \bar{x} = \frac{\sum_{i=1}^6 \hat{y}_i \cdot f_i}{n} = \frac{300}{28} = 10.71$$

$$\text{Vairanza } s^2 = \frac{\sum_{i=1}^6 (\hat{y}_i - \bar{x})^2 \cdot f_i}{n-1} = \frac{1060.71}{27} = 39.28$$

$$\text{Desviación } s = \sqrt{s^2} = 6.26$$

Intervalo de confianza

$N < 30$ , usamos distribución t-student.

$$\bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Por tablas  $t_{\alpha/2} = t_{0.05} = 1.703$ , con  $\nu = n - 1 = 27$  grados de libertad.

Por tanto

$$10.71 - 1.703 \frac{6.26}{\sqrt{28}} \leq \mu \leq 10.71 + 1.703 \frac{6.26}{\sqrt{28}}$$

$$8.69 \leq \mu \leq 12.72$$

Por tanto estamos un 90% seguros de que la media se va a encontrar entre 8.69 y 12.72

2. Los suponga que la media es 0.4, 0.3, 0.2 y 0.1, respectivamente, de que 0,1,2 o 3 fallas de energía eléctrica afecten cierta subdivisión en cualquier año dado. Encuentre la media y la varianza de la variable aleatoria X que representa el número de fallas de energía.

$$\bar{x} = \sum x_i \cdot p_i = 0.4 \cdot 0 + 0.3 \cdot 1 + 0.2 \cdot 2 + 0.1 \cdot 3 =$$

Se espera que en promedio 1 falla afecte cierta subdivisión cada año.

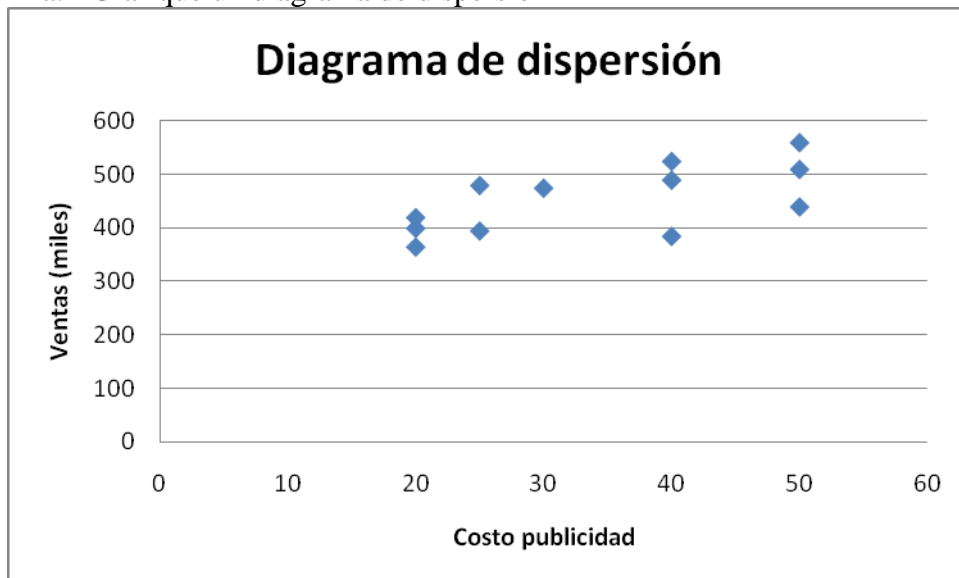
$$s^2 = \sum (x_i - \bar{x})^2 \cdot p_i = (0-1)^2 \cdot 0.4 + (1-1)^2 \cdot 0.3 + (2-1)^2 \cdot 0.2 + (3-1)^2 \cdot 0.1 = 1$$

Varianza = 1

3. Un comerciante al menudeo lleva a cabo un estudio para determinar la relación entre los gastos semanales de publicidad y de las ventas. Se registran los siguientes datos:

Costo de Publicidad(en orden de mil)	Ventas (en orden de mil)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

- a. Grafique un diagrama de dispersión



- b. Encuentre la ecuación de la línea de regresión para predecir las ventas semanales a partir de los gastos de publicidad

Dato No	Costo de Publicidad(en orden de mil) X	Ventas (en orden de mil) Y	X*Y	x^2
1	40	385	15400	1600
2	20	400	8000	400
3	25	395	9875	625
4	20	365	7300	400
5	30	475	14250	900
6	50	440	22000	2500

7	40	490	19600	1600
8	20	420	8400	400
9	50	560	28000	2500
10	40	525	21000	1600
11	25	480	12000	625
12	50	510	25500	2500
<b>Total</b>	<b>410</b>	<b>5445</b>	<b>191325</b>	<b>15650</b>

$$b = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{12 \cdot 191325 - 410 \cdot 5445}{12 \cdot 15650 - (410)^2} = 3.2208$$

$$a = \frac{\sum y_i}{n} - b \cdot \frac{\sum x_i}{n} = \frac{5445}{28} - 3.2208 \cdot \frac{410}{28} = 147.30$$

Por tanto la ecuación que estima las ventas (y) en función de la publicidad (x) es

$$\hat{y} = a + bx = 147.30 + 3.2208x$$

c. Estime las ventas semanales cuando los costos de publicidad son 35.000

$$\hat{y} = 147.30 + 3.2208 \cdot 35 = 260.03$$

El estimado en ventas es 260300

4. Un criminólogo realizó una investigación para determinar si la incidencia de ciertos tipos de crímenes varían de una parte a otra en una ciudad grande. Los crímenes particulares de interés son asalto, robo, hurto y homicidio. La siguiente tabla muestra el número de crímenes cometidos en cuatro áreas de la ciudad durante el año pasado.

Distrito	Asalto	Robo	Hurto	Homicidio
1	162	118	451	18
2	310	196	996	25
3	258	193	458	10
4	280	175	390	19

¿Podemos concluir a partir de estos datos con un nivel de significancia de 0.01 que la ocurrencia de estos tipos de crímenes es dependiente del distrito de la ciudad?

Realizamos una prueba de independencia

Distrito	Asalto		Robo		Hurto		Homicidio		Total
1	162	186,37	118	125,84	451	423,49	18	13,286031	749
2	310	379,96	196	256,56	996	863,38	25	27,0864745	1527
3	258	228,67	193	154,41	458	519,61	10	16,3015521	919
4	280	214,98	175	145,17	390	488,51	19	15,3259424	864
Total	1010		682		2295		72		4059

La nueva columna es el valor esperado de cada celda. Para ello, multiplicamos el valor del total de la fila por el total de la columna y lo dividimos entre el total de crímenes.

Ejemplo  $\frac{749 \cdot 1010}{4059} = 186.37$

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} = 124.52$$

$$\alpha = 0.001$$

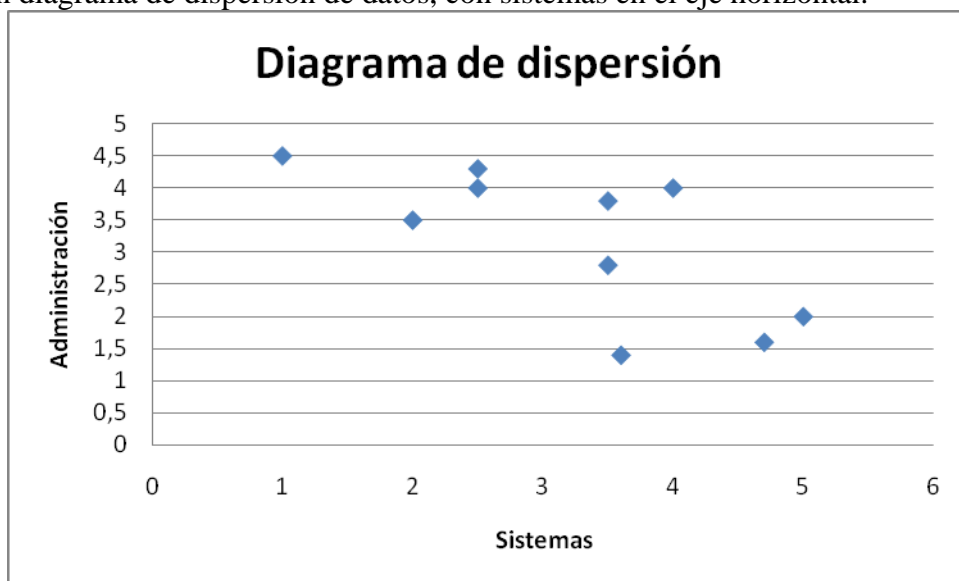
Región crítica  $\chi^2 > 27.877$  con  $\nu = (r - 1)(c - 1) = (4 - 1)(4 - 1) = 9$  grados de libertad

Por tanto cae en zona de rechazo. Por tanto los crímenes dependen del distrito.

5. A continuación se presenta las calificaciones de los 10 primeros estudiantes de la lista de estadística de Ingeniería de Sistemas y de Administración de Empresas

Calificac. Estd. Sistemas	Calificac. Estd. Admon de Empresas
3.5	3.8
4.0	4.0
2.5	4.3
3.5	2.8
4.7	1.6
5.0	2.0
3.6	1.4
2.0	3.5
2.5	4.0
1.0	4.5

Trace un diagrama de dispersión de datos, con sistemas en el eje horizontal.



a. ¿ Parece haber alguna relación entre las calificaciones de Sistemas y Administración ?. Explique su respuesta.

Aparentemente a medida que aumenta la nota para los de sistemas, disminuye la nota para los de administración. Es como si fuesen inversamente proporcionales las dos variables

b. Calcule e interprete la covarianza de la muestra de los datos.

Dato No	Calificac. Estd. Sistemas	$(x_i - \bar{x})$	Calificac. Estd. Admon de Empresas	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
1	3,5	0,27	3,8	0,61	0,1647
2	4	0,77	4	0,81	0,6237
3	2,5	-0,73	4,3	1,11	-0,8103
4	3,5	0,27	2,8	-0,39	-0,1053
5	4,7	1,47	1,6	-1,59	-2,3373
6	5	1,77	2	-1,19	-2,1063

7	3,6	0,37	1,4	-1,79	-0,6623
8	2	-1,23	3,5	0,31	-0,3813
9	2,5	-0,73	4	0,81	-0,5913
10	1	-2,23	4,5	1,31	-2,9213
<b>Total</b>	<b>32,3</b>		<b>31,9</b>		<b>-9,127</b>
Media	$\bar{x} = 3.23$		$\bar{y} = 3.19$		

La covarianza es  $\sigma_{XY} = \sum (x_i - \bar{x})(y_i - \bar{y}) = -9.12$

c. Calcule el coeficiente de correlación de la muestra de los datos. ¿Qué nos dice este valor acerca de la muestra de los datos ?.

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9.12}{1.2347 \cdot 1.1561} = -6.38$$

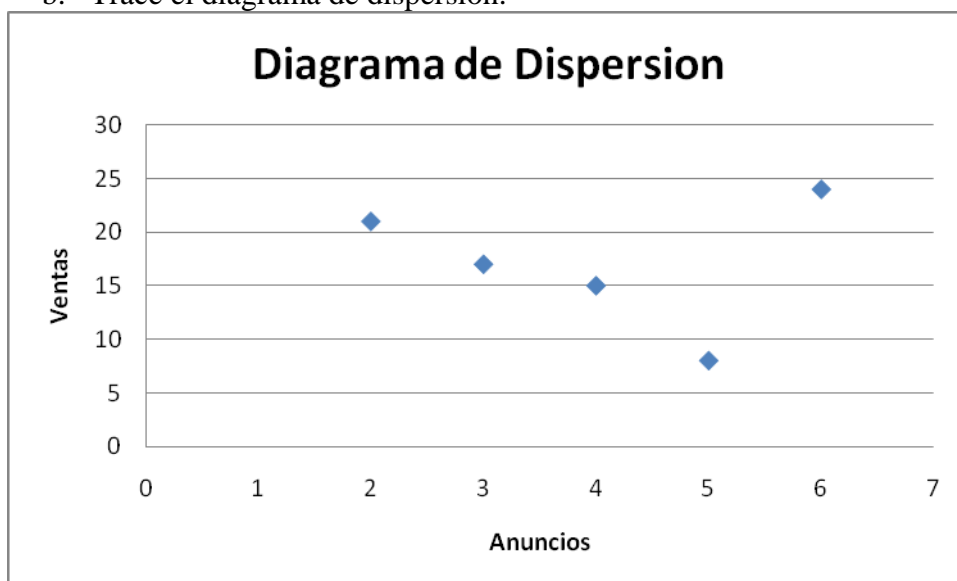
6. Una empresa comercial tiene establecimientos en varias zonas metropolitanas. La gerente general planea lanzar al aire un anuncio por televisión en algunas estaciones locales, al menos dos veces antes de realizar una venta gigante que ha de empezar el sábado y terminar el domingo. Planea tener las cifras de las ventas de videocámaras del sábado y el domingo en las diversas tiendas y agruparlas en pares con el número de veces que apareció el comercial en la televisión. El objeto fundamental de la investigación es determinar si existe alguna relación entre el número de veces que se transmitió el anuncio y las ventas de cámara de video. Los pares de datos son:

Localización de La televisora	Número de anuncios transmitidos	Ventas en sábado domingo (miles de dólares)
Bogotá	4	15
Cúcuta	2	8
Cali	5	21
Bucaramanga	6	24
Popayán	3	17

a. ¿Cuál es la variable independiente?

Número de anuncios transmitidos

b. Trace el diagrama de dispersión.



c. Calcule el coeficiente de correlación

	Anuncios	Ventas	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
Bogotá	4	15	0	-2	0
Cúcuta	5	8	1	-9	-9
Cali	2	21	-2	4	-8
Bucaramanga	6	24	2	7	14
Popayán	3	17	-1	0	0
<b>Total</b>	<b>20</b>	<b>85</b>			<b>-3</b>

<b>Media</b>	<b>4</b>	<b>17</b>			
<b>Desviación</b>	<b>1,58113883</b>	<b>6,12372436</b>			

$$\sigma_{XY} = \sum (x_i - \bar{x})(y_i - \bar{y}) = -3$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-3}{1.5811 \cdot 6.1237} = -0.3098$$

d. Interprete estas medidas estadísticas.

Por tanto, los datos tienen cierta tendencia de pendiente negativa, a más anuncios menos ventas. Sin embargo el valor no es concluyente ( No es cercano a -1)

7. Dentro de su entorno elabore un problema donde pueda aplicar los números índices donde trabaje por lo menos 8 años. ¿ Para qué utiliza los números índices en qué casos se pueden o se deben utilizar ?. ¿ Qué nos informa una relación de números índices ?

Un comerciante ha registrado las siguientes ventas anuales. Tomando como base el año 2000

<b>Año</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
<b>Ventas (\$)</b>	<b>200.000</b>	<b>250.000</b>	<b>200.000</b>	<b>190.000</b>	<b>220.000</b>

Cálculo de un índice de ventas

Año	Razón	Cambio de un decimal	Índice multiplicado x 100
2004	200.000/200.000	1.00	100
2005	250.000/200.000	1.25	125
2006	200.000/200.000	1.00	100
2007	190.000/200.000	0.95	95
2008	220.000/200.000	1.10	110

8. El diseñador de cierta marca de computadores afirma que el tiempo medio de los computadores es de 60000 horas. La desviación estándar es de 5000. Una empresa compró 48 computadores y halló que la duración media fue de 59500 horas. ¿Difiere la experiencia de la empresa de lo que afirma el diseñador? Use nivel de significancia de 0.05.

$$H_0: \mu = 60000$$

$$H_A: \mu < 60000$$

$$N = 48$$

$$\bar{x} = 59500$$

$$\sigma = 5000$$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{59500 - 60000}{\frac{5000}{\sqrt{48}}} = -0.6928$$

Si  $\alpha = 0.05$  el estadístico de comparación es  $Z_c = -1.645$

Como nuestro Z calculado es mayor que el Z de comparación, y cae en zona de aceptación, no existe evidencia estadística para afirmar que la media es menor a 60000 horas.

**CRITERIOS DE EVALUACIÓN:**

- Se evaluará el procedimiento y la claridad en el desarrollo de los problemas.
- La evaluación a distancia es individual.

**VALORACIÓN:**

- En la evaluación a distancia el primer punto tiene un valor de 1.0 el segundo y el tercer punto tiene un valor de 2.0 cada uno.