LISTA DE COTEJO DE INVESTIGACIÓN DOCUMENTAL

DOCENTE: PABLO PROMOTOR CAMPECHANO	ASIGNATURA: ESTADISTICA INFERENCIAL I	
DATOS GENERALES DEL PROCESO DE EVALUACIÓN		
NOMBRE DEL ALUMNO: MARTINEZ BARCENAS EMMANUEL	UNIDAD: 5	
PERIODO: AGOSTO-DICIEMBRE 2024	GRUPO: 301 C	FECHA DE ENTREGA: 8/12/2024
INSTRUCCIONES		
Revisar las actividades que se solicitan y marque en los apartados “SI” cuando la evidencia se cumple; en caso contrario marque “NO”. En la columna “OBSERVACIONES” indicaciones que puedan ayudar al alumno a saber cuáles son las condiciones no cumplidas, si fuese necesario.		

VALOR DEL REACTIVO	CARACTERÍSTICA A CUMPLIR (REACTIVO)	CUMPLE		OBSERVACIONES
		SI	NO	
5%	**PRESENTACIÓN:** la investigación cumple con los requisitos de: **a.** Buena presentación **b.** letra legible **c.** Limpieza y orden **d.** Ortografía (El documento es redactado de forma correcta sin faltas de ortografía)	√		
5%	Maneja el lenguaje técnico apropiado y presenta en todo el documento coherencia y secuencia entre párrafos	√		
5%	**INTRODUCCIÓN**: Da una idea clara y objetiva de lo que tratara el tema (motivando al lector a continuar con su lectura y revisión), fundamentando con las referencias bibliográficas que se utilizaron para su redacción.	√		
25%	**CONTENIDO**: Maneja un lenguaje técnico apropiado y presenta en todo el documento coherencia, secuencia entre párrafo, es digerible a todo público y presenta una metodología **COHERENCIA Y COHESIÓN:** Maneja un lenguaje técnico apropiado y presenta en todo el documento coherencia, secuencia entre párrafo y es digerible a todo público coherente.	√		
5%	**Conclusiones**: Las conclusiones son claras y acordes con el objetivo esperado.		√	
5%	**Responsabilidad**: Entregó la investigación documental en la fecha señalada.	√		
50%	CALIFICACIÓN	45%		

INSTITUTO TECNOLÓGICO SUPERIOR DE SAN ANDRÉS TUXTLA

Estadística

ALUMNOS:

EMMANUEL MARTINEZ BARCENAS

MARÍA FERNANDA TON LÓPEZ

JESÚS VELASCO CATEMAXCA

GRUPO:

301 – C

ALUMNO:

EMMANUEL MARTINEZ BARCENAS

MARÍA FERNANDA TON LÓPEZ

JESÚS VELASCO CATEMAXCA

GRUPO:

301 – C

ASIGNATURA:

ESTADISTICA INFERENCIAL 1

DOCENTE:

PABLO PROMOTOR CAMPECHANO

TEMA:

INVESTIGACION UNIDAD 5

FECHA:

08/12/24

INDICE

INTRODUCCIÓN

Frecuentemente estamos interesados en explorar los efectos de distintos factores sobre una variable de interés herramental más elemental y usada que ocupamos los economistas para esto es un método estadístico llamado análisis de regresión, el cual se usa para estudiar el efecto de una o más variable independientes sobre una variable dependiente. Una variable dependiente es una variable de resultado que queremos explicar usando otras variables. Las variables independientes son las variables que usamos para "explicar la variación en la variable dependiente (también son llamadas variables explicativas). Uno de los aspectos más relevantes de la Estadística es el análisis de la relación o dependencia entre variables. Frecuentemente resulta de interés conocer el efecto que una o varias variables pueden causar sobre otra, e incluso predecir en mayor o menor grado valores en una variable a partir de otra. Por ejemplo, supongamos que la altura de los padres influye significativamente en la de los hijos. Podríamos estar interesados en estimar la altura media de los hijos cuyos padres presentan una determinada estatura.

5.1 REGRESIÓN LINEAL SIMPLE

Si sabemos que existe una relación entre una variable denominada dependiente y otras denominadas independientes (como por ejemplo las existentes entre: la experiencia profesional de los trabajadores y sus respectivos sueldos, las estaturas y pesos de personas, la producción agraria y la cantidad de fertilizantes utilizados, etc.), puede darse el problema de que la dependiente asuma múltiples valores para una combinación de valores de las independientes.

La dependencia a la que hacemos referencia es relacional matemática y no necesariamente de causalidad. Así, para un mismo número de unidades producidas, pueden existir niveles de costo, que varían empresa a empresa.

Si se da ese tipo de relaciones, se suele recurrir a los estudios de regresión en los cuales se obtiene una nueva relación, pero de un tipo especial denominado función, en la cual la variable independiente se asocia con un indicador de tendencia central de la variable dependiente. Cabe recordar que, en términos generales, una función es un tipo de relación en la cual para cadavalor de la variable independiente le corresponde uno y sólo un valor de la variable dependiente.

La Regresión y la correlación son dos técnicas estadísticas que se pueden utilizar para solucionar problemas comunes en los negocios.

Muchos estudios se basan en la creencia de que es posible identificar y cuantificar alguna Relación Funcional entre dos o más variables, donde una variable depende de la otra variable.

Se puede decir que Y depende de X, en donde Y y X son dos variables cualquiera en un modelo de Regresión Simple.

"Y es una función de X"

Y = f(X)

Como Y depende de X,

Y es la variable dependiente, y

X es la variable independiente.

En el Modelo de Regresión es muy importante identificar cuál es la variable dependiente y cuál es la variable independiente.

En el Modelo de Regresión Simple se establece que Y es una función de sólo una variable independiente, razón por la cual se le denomina también Regresión Di variada porque sólo hay dos variables, una dependiente y otra independiente y se representa así:

Y = f (X)

"Y está regresando por X"

La variable dependiente es la variable que se desea explicar, predecir. También se le llama REGRESANDO o VARIABLE DE RESPUESTA.

La variable Independiente X se le denomina VARIABLE EXPLICATIVA ó REGRESOR y se le utiliza para EXPLICAR Y.

Ejemplo

Los datos de la siguiente tabla representan las estaturas (X, cm) y los pesos (Y, kg) de una muestra de 12 hombres adultos. Para cada estatura fijada previamente se observó el peso de una persona seleccionada de entre el grupo con dicha estatura, resultando:

X	152	155	152	155	157	152	157	165	162	178	183	178
Y	50	61.5	54.5	57.5	63.5	59	61	72	66	72	84	82

Con estos datos vamos a plantear una ecuación de regresión simple que nos permita pronosticar los pesos conociendo las tallas. Utilizaremos a = 0.05, y contrastaremos nuestra hipótesis con la prueba F.

DESARROLLO

- Representación matemática y gráfica de los datos:

Representación Matemática

	estatura	pesos		Regresión Lineal				I.C. para la media		I. C. individual	
datos	x	y	x ^2	y ^2	Xy	y est.	Residual	L. I.	L. S.	L. I.	L. S.
1	152	50	23104	2500	7600	56.43	-6.43	53.07	59.79	47.30	65.56
2	155	61.5	24025	3782.3	9532.5	59.03	2.47	56.09	61.97	50.05	68.02
3	152	54.5	23104	2970.3	8284	56.43	-1.93	53.07	59.79	47.30	65.56
4	155	57.5	24025	3306.3	8912.5	59.03	-1.53	56.09	61.97	50.05	68.02
5	157	63.5	24649	4032.3	9969.5	60.77	2.73	58.05	63.48	51.85	69.68
6	152	59	23104	3481	8968	56.43	2.57	53.07	59.79	47.30	65.56
7	157	61	24649	3721	9577	60.77	0.23	58.05	63.48	51.85	69.68
8	165	72	27225	5184	11880	67.71	4.29	65.17	70.24	58.85	76.57

9	162	66	26244	4356	10692	65.11	0.89	62.65	67.56	56.27	73.94
10	178	72	31684	5184	12816	78.99	-6.99	74.65	83.33	69.45	88.52
11	183	84	33489	7056	15372	83.32	0.68	78.01	88.64	73.31	93.34
12	178	82	31684	6724	14596	78.99	3.01	74.65	83.33	69.45	88.52

5.2 CALIDAD DEL AJUSTE EN REGRESIÓN LINEAL SIMPLE

En la sección anterior estudiamos pruebas de hipótesis para verificar que hay una relación significativa entre y; sin embargo, no hemos visto si tal relación permite hacer estimaciones con una precisión aceptable. Por ejemplo, es de interés saber qué tanta de la variabilidad presente en fue explicada por el modelo, además si se cumplen los supuestos de los residuos.

Coeficiente de determinación. Un primer criterio para evaluar la calidad del ajuste es observar la forma en que el modelo se ajustó a los datos. En el caso de la regresión lineal simple esto se distingue al observar si los puntos tienden a ajustarse razonablemente bien a la línea recta (véase la figura 1.3). Pero otro criterio más cuantitativo es el que proporciona el coeficiente de determinación, el cual está definido por:

$$R^2 = \frac{\text{Variabilidad explicada por el modelo}}{\text{Variabilidad total}} \longrightarrow R^2 = \frac{SC_R}{S_{yy}}$$

En general se interpreta como la proporción de la variabilidad en los datos (Y) que es explicada por el modelo. En el caso de los datos de la resistencia de la pulpa (tabla 1.1) tenemos

$$SC_R = \hat{B}_1 S_{xy} = (1{,}6242)(1478{,}0) = 2400{,}5676$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y\right)^2}{n} = 353{,}342 - \frac{2216^2}{14} = 2580{,}86$$

$$R^2 = \frac{SC_R}{S_{yy}} \quad = \frac{2\,400{,}56}{2\,580{,}86} = 0{,}93$$

Por lo tanto, podemos decir que 93% de la variación observada en la resistencia es explicada por el modelo (línea recta), lo cual nos dice que la calidad del ajuste es satisfactoria, y que, por ello la relación entre es descrita adecuadamente por una línea recta.

EJEMPLO

Ecuación lineal de la correlación entre la variable y y la variable x

Coeficiente de correlación lineal

$$y = ax + b$$

y	x	x^2	xy
3	100	10000	300
5	90	8100	450
9	80	6400	720
10	45	2025	450
20	50	2500	1000
21	50	2500	1050
24	60	3600	1440
24	40	1600	960
27	25	625	675
35	20	400	700
178	560	37750	7745

$n = 10$

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$a = \frac{10(7745) - (560)(178)}{10\,(37750) - \,(560)^2}$$

$$= \frac{77450 - 99680}{377500 - 313600}$$

$$= \frac{-22230}{63900} = -0.34788$$

Ecuación lineal de la correlación entre la variable ***y*** y la variable ***x***

Coeficiente de correlación lineal

y	x	x^2	xy
3	100	10000	300
5	90	8100	450
9	80	6400	720
10	45	2025	450
20	50	2500	1000
21	50	2500	1050
24	60	3600	1440
24	40	1600	960
27	25	625	675
35	20	400	700
178	560	37750	7745

$n = 10 \qquad a = -0.34788$

$$y = ax + b$$

$$b = \frac{\sum y - a\sum x}{n}$$

$$b = \frac{178 - (-0.34788)(560)}{10}$$

$$b = \frac{178 + 194.8128}{10}$$

$$b = \frac{372.8128}{10} = 37.28128$$

Ecuación lineal de la correlación entre la variable ***y*** y la variable ***x***

Coeficiente de correlación lineal

y	x	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
3	100	44	-14.8	1936	219.04	-651.2
5	90	34	-12.8	1156	163.84	-435.2
9	80	24	-8.8	576	77.44	-211.2
10	45	-11	-7.8	121	60.84	85.8
20	50	-6	2.2	36	4.84	-13.2
21	50	-6	3.2	36	10.24	-19.2
24	60	4	6.2	16	38.44	24.8
24	40	-16	6.2	256	38.44	-99.2
27	25	-31	9.2	961	84.64	-285.2
35	20	-36	17.2	1296	295.84	-619.2
178	560			6390	993.6	-2223

$$\bar{x} = \frac{\sum x}{n} = \frac{560}{10} = 56$$

$$\bar{y} = \frac{\sum y}{n} = \frac{178}{10} = 17.8$$

Ecuación lineal de la correlación entre la variable y y la variable x

Coeficiente de correlación lineal

$$y = ax + b$$

y	x	x^2	xy
3	100	10000	300
5	90	8100	450
9	80	6400	720
10	45	2025	450
20	50	2500	1000
21	50	2500	1050
24	60	3600	1440
24	40	1600	960
27	25	625	675
35	20	400	700
178	560	37750	7745

$$y = \boxed{-0.34788x + 37.28128}$$

$n = 10$ $\quad a = -0.34788 \quad b = 37.28128$

Ecuación lineal de la correlación entre la variable y y la variable x

Coeficiente de correlación lineal

$$\sum(x-\bar{x})(y-\bar{y}) = -2223$$

$$\sum(x-\bar{x})^2 = 6390$$

$$\sum(y-\bar{y})^2 = 993.6$$

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}$$

$$r = \frac{-2223}{\sqrt{6390}\sqrt{993.6}} = \frac{-2223}{(79.93747)(31.52142)}$$

$$r = -0.88223$$

5.3 ESTIMACIÓN Y PREDICCIÓN POR INTERVALO EN REGRESIÓN LINEAL

Una de las aplicaciones más importantes en un análisis de regresión es hacer estimaciones de la respuesta media para un valor dado de X. En el caso particular de la regresión lineal simple, sabemos que un estimador puntual de la respuesta media lo da la recta de regresión:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Además de esto, en ocasiones es de interés obtener una estimación por intervalos para a partir de cualquier valor de X, para lo cual aplicamos la siguiente ecuación:

$$\hat{y}_0 - t_{(\alpha/2,\ n-2)}\sqrt{CM_E\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \leq E(y\backslash x_0) \leq \hat{y}_0 + t_{(\alpha/2,\ n-2)}\sqrt{CM_E\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]} \quad (1.21)$$

EJEMPLO

Ejemplo:

El gerente de personal de una empresa intuye que quizás haya relación entre el ausentismo en días (Y) y la edad en años (X) por lo que quiere tomar la edad de un trabajador para desarrollar un modelo de predicción de días de ausencia durante un año laboral. Se seleccionó una muestra aleatoria de 10 trabajadores con los resultados que se presentan a continuación:

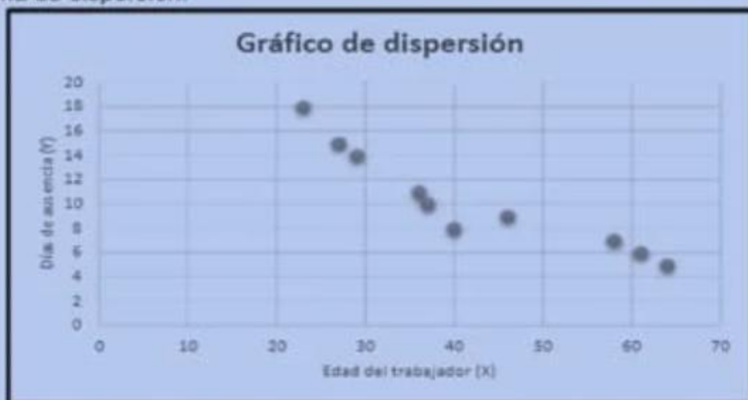
Edad en años (X)	Ausentismo en días (Y)
27	15
61	06
37	10
23	18
46	09
58	07
29	14
36	11
64	05
40	08

Solución

a) El diagrama de dispersión es una serie de puntos donde se intersecta el valor de la variable "x" y el valor de la variable "y" que se ubican en el plano cartesiano.

Así tenemos por ejemplo que una persona de 64 años, se ha ausentado 5 días en el año. Esto se representa por el punto (64,5), una persona de 23 años se ha ausentado 18 días en el año y así sucesivamente.

Una vez ubicado todos los puntos en el I cuadrante del plano obtenemos el siguiente diagrama de dispersión:



b) Para encontrar la función de regresión muestral tenemos que encontrar todas las sumas de cuadrados de las variables en cuestión. Entonces procederemos a completar la tabla dada y a continuación sustituimos en las fórmulas correspondientes:

Trabajadores	Edad en años (X)	Ausentismo en días (Y)	X^2	Y^2	XY
1	27	15	729	225	405
2	61	6	3721	36	366
3	37	10	1369	100	370
4	23	18	529	324	414
5	46	9	2116	81	414
6	58	7	3364	49	406
7	29	14	841	196	406
8	36	11	1296	121	396
9	64	5	4096	25	320
10	40	8	1600	64	320
Totales	**421**	**103**	**19661**	**1221**	**3817**

$$b_1 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

$$b_1 = \frac{3{,}817 - \frac{(421)(103)}{10}}{19{,}661 - \frac{(421)^2}{10}} = \frac{3{,}817 - 4{,}336.30}{19{,}661 - 17{,}724.10} = \frac{-519.30}{1{,}936.90} = -0.268$$

$$b_0 = \frac{\sum Y_i - b_1 \sum X_i}{n}$$

$$b_0 = \frac{103 - (-0.268)(421)}{10} = \frac{103 + 112.87}{10} = \frac{215.87}{10} = 21.587$$

CONCLUSIÓN

La finalidad de la inferencia estadística es obtener información sobre características desconocidas de las poblaciones generalmente cuantificadas por parámetros a partir de características conocidas de las muestras cuantificadas por estadísticos. Incluir poblaciones o inmensos datos de individuos en la investigación suele ser impracticable, y por ello se suele trabajar con grupos pequeños generalizando los resultados mediante las técnicas de Estadística Inferencia.

BIBLIOGRAFIA

http://dta.utalca.cl/estadistica/ejercicios/interpretar/Metodos/propuesto%20NOPARAM.pdf

http://www.dm.uba.ar/materias/estadistica_Q/2010/2/C010%20Metodos%20no%20param%20(una%20muestra).pdf

LISTA DE COTEJO DE PROBLEMARIO

<table>
<tr><td colspan="2">DOCENTE: ING. PABLO PROMOTOR CAMPECHANO</td><td colspan="3">ASIGNATURA: ESTADISTICA INFERENCIAL I</td></tr>
<tr><td colspan="5">DATOS GENERALES DEL PROCESO DE EVALUACIÓN</td></tr>
<tr><td colspan="3">NOMBRE DEL ALUMNO (A): MARTINEZ BARCENAS EMMANUEL</td><td colspan="2">Problemario de la Unidad: 5</td></tr>
<tr><td>PERIODO: AGOSTO-DICIEMBRE 2024</td><td>GRUPO: 301 C</td><td colspan="3">FECHA DE ENTREGA: 8/12/2024</td></tr>
<tr><td colspan="5">INSTRUCCIONES</td></tr>
<tr><td colspan="5">Revisar las actividades que se solicitan y marque en los apartados “SI” cuando la evidencia se cumple; en caso contrario marque “NO”. En la columna “OBSERVACIONES” indicaciones que puedan ayudar al alumno a saber cuáles son las condiciones no cumplidas, si fuese necesario.</td></tr>
</table>

<table>
<tr><th rowspan="2">VALOR DEL REACTIVO</th><th rowspan="2">CARACTERÍSTICA A CUMPLIR (REACTIVO)</th><th colspan="2">CUMPLE</th><th>OBSERVACIONES</th></tr>
<tr><th>SI</th><th>NO</th><th></th></tr>
<tr><td>5 %</td><td>PRESENTACIÓN: El trabajo cumple con los requisitos de
a. Buena presentación
b. No tiene faltas de ortografía
c. Ordenado y limpio</td><td>√</td><td></td><td></td></tr>
<tr><td>5 %</td><td>FORMATO DE ENTREGA: Hoja de presentación (asignatura, unidad, tema de estudio, docente, fecha, nombre del alumno</td><td></td><td>√</td><td></td></tr>
<tr><td>30 %</td><td>DESARROLLO DE EJERCICIOS: Identifica los principios, leyes, normas e incluso técnicas y metodologías apropiadas, si el ejercicio lo permite, debe de presentar datos, fórmula, sustitución y resultado.</td><td>√</td><td></td><td></td></tr>
<tr><td>5 %</td><td>RESULTADO: El alumno llega al resultado correcto, con sus respectivas unidades</td><td>√</td><td></td><td></td></tr>
<tr><td>5 %</td><td>RESPONSABILIDAD: Entregó el problemario en la fecha señalada.</td><td>√</td><td></td><td></td></tr>
<tr><td>50 %</td><td>CALIFICACIÓN</td><td colspan="3">45%</td></tr>
</table>

unidad v
regresión lineal
simple

Un comerciante minorista realizó un estudio para determinar la relación que hay entre los gastos semanales de publicidad y las ventas

a) calcular la ecuación de las rectas de regresión
b) grafique la recta en un diagrama de dispersión
c) determina las ventas (y) si el costo de publicidad (x) fuera de 48, 51 y 60

	costo publicidad(x)	Ventas (y)	xy	x^2
1	42	385	16170	1764
2	21	400	8400	441
3	25	396	9875	625
4	20	365	7300	400
5	30	475	14250	900
6	50	440	22000	2500
7	45	490	22050	2025
8	22	420	9240	484
9	50	560	28000	2500
10	46	525	24150	2116
	351	4455	161435	13755

a) $b_1 = \dfrac{(n\sum xy) - (\sum x)(\sum y)}{(n\sum x^2) - (\sum x)^2}$

$b_0 = \dfrac{(\sum y) - b_1(\sum x)}{n}$

$\hat{y} = b_0 + b_1 x$

$b_0 = \dfrac{4455 - 3.52(351)}{10}$

$b_0 = \dfrac{3219.48}{10}$

$\boxed{b_0 = 321.6}$

$b_1 = \dfrac{10(161435) - (351)(4455)}{10(13755) - (351)^2}$

$\boxed{\hat{Y} = 321.6 + 3.53X}$

$b_1 = \dfrac{50645}{14349}$

$\boxed{b_1 = 3.529}$

c) $\hat{y} = 321.6 + 3.53(48) = 491.04$

$\hat{y} = 321.6 + 3.53(51) = 501.63$

$\hat{y} = 321.6 + 3.53(60) = 533.4$

	X	Y	XY	X^2	$\hat{y}$	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(y-\hat{y})$	$(y-\hat{y})^2$
1	42	385	16170	1764	469.84	6.9	-60.5	47.61	3660.25	-84.84	7197.82
2	21	400	8400	441	395.72	-14.1	-45.5	198.81	2070.25	4.28	18.31
3	25	395	9875	625	409.83	-10.1	-60.5	102.01	3660.25	-14.83	219.92
4	20	365	7300	400	392.19	-15.1	-80.5	228.01	6480.25	-29.19	739.29
5	30	475	14250	900	427.48	-5.1	29.5	26.01	870.25	47.52	2258.15
6	50	440	22000	2500	498.07	14.9	-5.5	222.01	30.25	-58.07	3372.12
7	45	490	22050	2025	480.43	9.9	44.5	98.01	1980.25	9.57	91.58
8	22	420	9240	484	399.25	-13.1	-25.5	171.61	650.25	20.75	430.56
9	50	560	28000	2500	498.07	14.9	114.5	222.01	13110.25	61.93	3835.32
10	46	525	24150	2116	483.95	10.9	79.5	118.81	6320.25	41.05	1685.10
		4455	161435	13755				1434.9	37722.5		19941.38

$\hat{y} = 321.59 + 3.53(x)$

$\bar{x} = \frac{351}{10} = 35.1$

$\bar{y} = \frac{4455}{10} = 445.5$

$S_{xx} = \sum_{i=1}^{n} (x-\bar{x})^2$

$S_{yy} = \sum_{i=1}^{n} (y-\bar{y})^2$

$S_{xy} = \sum_{i=1}^{n} (x-\bar{x})(y-\bar{y})$

$SCE = \sum_{i=1}^{n} (y-\hat{y})^2$

$S^2 = \frac{SCE}{n-2}$

$S = \sqrt{S^2}$

Coeficiente de correlación $(r) = b_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$

$(r) = (3.529)\sqrt{\frac{1434.9}{37722.5}} = 0.68$

Coeficiente de determinación $(r^2) = \frac{S_{xy}^2}{S_{xx}\,S_{yy}}$

$(r^2) = $ [illegible]

$(r^2) = $ [illegible]

$(r^2) = 0.47$

ploblemario

unidad 5

1 De una determinada empresa se conocen los siguientes datos, referidos al volumen de ventas (millones de pesos) y al gasto de publicidad (miles de pesos) de los últimos 6 años

x	y	xy	x^2	$\hat{y}$	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$	$(y-\hat{y})^2$
16	10	160	256	10.433	-34	-12.33	1156	152.028	419.22	0.187
32	15	480	1024	16.033	-18	-7.33	324	53.728	131.94	1.067
48	25	1200	2304	21.633	-2	2.67	4	7.128	-5.34	11.336
60	22	1320	3600	25.833	10	-0.33	100	0.108	-3.3	14.691
64	30	1920	4096	27.233	14	7.67	196	58.828	107.38	7.656
80	32	2560	6400	32.833	30	9.67	900	93.508	290.1	0.693
T=300	134	7640	17,680				2680	365.328	940	35.63

$$b_1 = \frac{(n\sum xy) - (\sum x)(\sum y)}{(n\sum x^2) - (\sum x)^2}$$

$$b_1 = \frac{6(7640) - (300)(134)}{6(17680) - (300)^2}$$

$$b_1 = \frac{5640}{16{,}080}$$

$$b_1 = 0.350$$

$$b_0 = \frac{(\sum y) - b_1(\sum x)}{n}$$

$$b_0 = \frac{134 - (0.350)(300)}{6}$$

$$b_0 = \frac{29}{6}$$

$$b_0 = 4.833$$

$$\hat{y} = b_0 + b_1(x)$$

$$\hat{y} = 4.833 + 0.350\,x$$

$$\hat{y} = 4.833 + 0.350(75)$$

$$\hat{y} = 31.083$$

$$\bar{x} = \frac{300}{6} = 50$$

$$\bar{y} = \frac{134}{6} = 22.33$$

Scribe

$$S_{xx} = \sum_{i=1}^{n} (x-\bar{x})^2 = 2680$$

$$S_{yy} = \sum_{i=1}^{n} (y-\bar{y})^2 = 365.328$$

$$S_{xy} = \sum_{i=1}^{n} (x-\bar{x})(y-\bar{y}) = 940$$

$$SCE = \sum_{i=1}^{n} (y-\hat{y})^2 = 35.63$$

$$S^2 = \frac{SCE}{n-2} = \frac{35.63}{6-2} = \frac{35.63}{4} = 8.907$$

$$S = \sqrt{S^2} = 2.98$$

b_1	0.350
b_0	4.833
S_{xx}	2680
S_{yy}	365.328
S_{xy}	940
SCE	35.63
S^2	7.126
S	2.66

Coeficiente de correlación $= (r) = b_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$

$$(r) = (0.350)\sqrt{\frac{2680}{365.328}}$$

$$(r) = 0.94$$

Coeficiente de determinación $= (r^2) = \frac{S_{xy}^2}{S_{xx}\, S_{yy}}$

$$(r^2) = \frac{(940)^2}{(2680)(365.328)}$$

$$(r^2) = \frac{883{,}600}{979{,}079.04}$$

$$(r^2) = 0.90$$

$$S_{xx} = \sum_{i=1}^{n} (x-\bar{x})^2 = 82,304$$

$$S_{yy} = \sum_{i=1}^{n} (y-\bar{y})^2 = 98.24$$

$$S_{xy} = \sum_{i=1}^{n} (x-\bar{x})(y-\bar{y}) = 2061.44$$

$$SCE = \sum_{i=1}^{n} (y-\hat{y})^2 = 37.401$$

$$S^2 = \frac{SCE}{n-2} = \frac{37.401}{8} = 4.675$$

$$S = \sqrt{S^2} = 2.162$$

b_1	0.036
b_0	-2.032
S_{xx}	82,304
S_{yy}	98.24
S_{xy}	2061.44
SCE	37.401
S^2	4.675
S	2.162

Coeficiente de correlación $= (r) = b_1 \sqrt{\frac{S_{xx}}{S_{yy}}}$

$$(r) = (0.036)\sqrt{\frac{82,304}{98.24}}$$

$$(r) = 1.042$$

Coeficiente de determinación $= (r^2) = \frac{S_{xy}^2}{S_{xx}\,S_{yy}}$

$$(r^2) = \frac{(2061.44)^2}{(82,304)(98.24)}$$

$$(r^2) = \frac{4249534.874}{8085544.96}$$

$$(r^2) = 0.525$$

Problemario

Un estudio de mercado trata de averigoar si es efectiva la propaganda televisada de un producto que salio a la venta con relacion al tiempo de publicidad (en horas / semana). Se recopilaron datos a a partir de la segunda semana de iniciada la publicidad resultando el cuadro que si: No se pudo recopilar datos de la cuarta semana:

Semana	2	3	4	5	6	7	---
venta de productos	300	310	–	320	350	420	
Tiempo de Propaganda	20	25	22	28	36	40	

a) ¿Es efectiva la publicidad del producto?

b) ¿En cuanto se estimaria las ventas para la semana 4?

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	20	300	6000	400	900 000
2	25	310	7750	625	96 100
3	28	320	8960	784	102 400
4	36	350	12 600	1296	122 500
5	40	420	16 800	1600	176 400
6	149	1,700	52 110	4705	587 400

Tiempo de propaganda (x): variable independiente

Venta del producto e [illegible] (y): variable independiente

a) $y = a + bx$ donde $b = \frac{S_{xy}}{S_x^2} = \frac{290}{52.96} = 5.476$

$a = \bar{y} - b\bar{x} = 340 - 5.476\,(29.8)$

$= 176.82$

$y = 176.82 + 5.476\,x$

$r = \frac{S_{xy}}{S_x S_y} = \frac{290}{(7.2774)(43.389)} = \frac{290}{(7.27749)(43.389)}$

$= 0.92 \cong 92\,\%$

b) Hallamos "y" para $x = 22$

$y = 176.82 + 5.476\,(22) = 297.29 \cong \$\,297$

2. Un editor tomo una muestra de 7 libros anotando el precio y el numero de paginas respectivas.
Obteniendo los siguientes datos:

Nº de pagina	630	550	400	250	370	320	610
Precio	10	8	7	4	6	6	9

a) Determine una funcion lineal entre el precio y el numero de paginas con el fin de predecir precios.
b) Estimar el precio de un libro de 300 paginas. Si a este libro se le incrementan 20 paginas en una segunda edicion. ¿En cuanto se incrementaria su precio?
c) ¿Cuantas paginas deberia tener un libro cuyo precio se estima en $ 12.27?

Numero de paginas (x): variable Independiente
precio en $(y): variable Independiente

	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	630	10	6300	396 900	100
2	550	8	4400	302 500	64
3	400	7	2800	160 000	49
4	250	4	1000	62 500	16
5	370	6	2220	136 900	36
6	320	6	1920	102 400	36
7	610	9	5490	372 100	81

$x_i = 3130$
$y_i = 50$
$x_i \cdot y_i = 24130$
$x_i^2 = 1\,533,300$
$y_i^2 = 382$

$\bar{x} = \frac{\sum xi}{n} = \frac{3130}{7} = 447.1429$

$\bar{y} = \frac{\sum yi}{n} = \frac{1533300}{7} = 7.1429$

$= 138.2248$

$Sx^2 = \frac{\sum yi^2}{n} - (\bar{y})^2 = \frac{382}{7} - (7.1429)^2 = 3.5504$

$\rightarrow Sx = 1.8843$

$Con_{xy} = S_{xy} = \frac{\sum xi\, yi}{n} - \bar{x}\,(\bar{y}) = \frac{24130}{7} - 447.1427$

$(7.1429) = 253.2458$

a) $y = a + bx$ donde $b = \frac{S_{xy}}{Sx^2} = \frac{253.2458}{19106.0842}$

$= 0.01325$

$a = \bar{y} - \bar{b}x = 7.1429 - 0.01325\,(447.1429) = 1.22$

$y = 1.22 + 0.013x$

b) $y = 1.22 + 0.013x \rightarrow 1.22 + 0.013\,(300) = \$\,5.12$

El b representa el cambio en Y por cada unidad de cambio en x

El Incremento sera : $0.013 \times 20 = \$\,0.26$

c) $y = 1.22 + 0.013x \rightarrow 1.22 + 0.013(x) = 12.27$

$0.013x = 11.05 \rightarrow x = 850$ paginas