A menudo, en la práctica, se requiere resolver problemas que incluyen conjuntos de variables, cuando se sabe que existe alguna relación inherente entre ellas, esa relación se puede encontrar a partir de la información experimental. El aspecto estadístico se convierte entonces en lograr la mejor estimación de la relación entre las variables.

En las aplicaciones hay una distinción clara entre las variables y su rol en el proceso experimental. Muy a menudo existe una sola variable dependiente o de **respuesta Y**, que no se controla en el experimento. Esta respuesta depende de una o más variables de regresión **independientes**, $x_1, x_2 \ldots . x_{k,}$ que se miden con un error insignificante y a menudo realmente se controlan en el experimento y por lo tanto **no son variables aleatorias,** y por tanto no tienen propiedades de distribución.

La relación que se ajusta a un conjunto de datos experimentales se caracteriza por una ecuación de predicción que se denomina ecuación de regresión.

ECUACIÓN DE PREDICCIÓN = ECUACIÓN DE REGRESIÓN

LEOPOLDO VIVEROS ROSAS

En el caso de una sola Y y una sola x, la situación se convierte en una regresión de Y sobre x. Para k variables independientes se habla de una variable de respuesta Y sobre , $x_1, x_2 .... x_{k,}$.

Denotemos una muestra aleatoria de tamaño n con el conjunto $\{(x_i, y_i); i=1,2,\ldots,n\}$. Si se toman muestras adicionales mediante el uso de exactamente los mismos valores de x, debemos esperar que varíen los valores de y. De aquí el valor y_i en el par ordenado (x_i, y_i) es un valor de alguna variable aleatoria Y_i. Por conveniencia definimos Y/x como la variable aleatoria Y que corresponde a un valor fijo x y denotamos su media y su varianza como $\mu_{Y/x}$ y su varianza como $\sigma^2{}_{Y/x}$. Es claro que si x= xi, el símbolo Y/xi representa la variable aleatoria Yi con media $\mu_{Y/xi}$ y varianza $\sigma^2{}_{Y/xi}$.

El término regresión lineal implica que $\mu_{Y/x}$ se relaciona linealmente con x mediante la ecuación de regresión de población

$$\mu_{Y/x} = \alpha + \beta\chi$$

Donde los coeficientes de regresión α y β son parámetros a estimar a partir de los datos muestrales. Al denotar sus estimaciones como a y b respectivamente, podemos estimar $\mu_{Y/x}$ con $\hat{y}$ a partir de la regresión de la muestra o l línea de regresión ajustada

$$\hat{y} = a + bx$$

Donde las estimaciones a y b representan la intersección y la pendiente respectivamente.

Si postulamos que todas las medias $\mu_{Y/xi}$ caen en una línea recta , cada Yi se puede describir como un modelo de regresión lineal simple

$$Y_i = \mu_{Y/x_i} + E_i = \alpha + \beta x_i + E_i$$

Donde el error aleatorio Ei, el error del modelo, necesariamente debe tener una media de cero. Cada observación (x_i, y_i) en nuestra muestra satisface la ecuación

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Donde εi es el valor que toma E_i cuando Y_i toma el valor y_i. La ecuación anterior se puede ver como el modelo para una sola observación y_i .

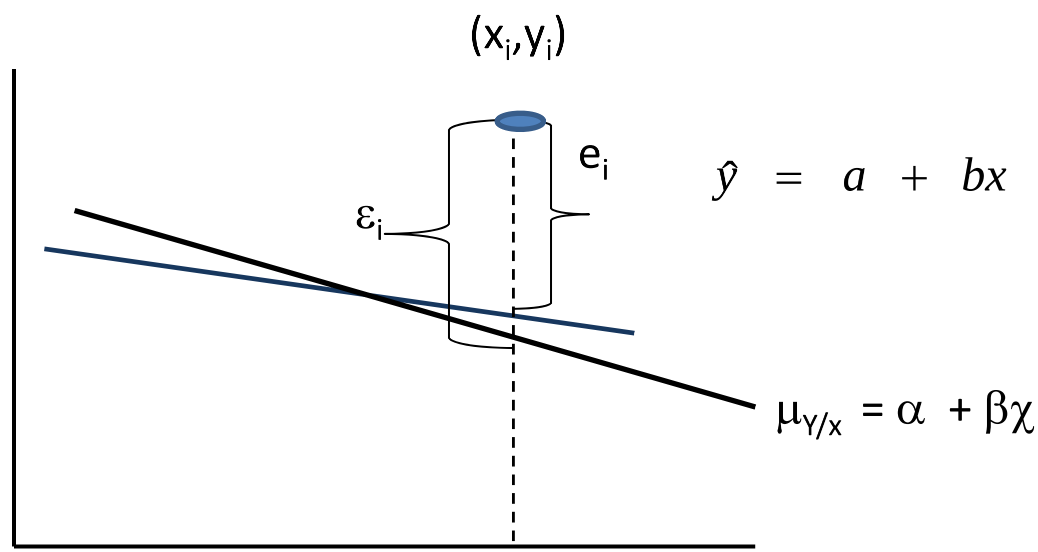
De manera similar, con el uso de la regresión estimada o ajustada

$$\hat{y} = a + bx$$

Cada par de observaciones satisface la relación

$$\hat{y}_i = a + bx_i + e_i$$

Donde $e_i = y_i - \hat{y}_i$ se denomina residuo y describe el error en el ajuste del modelo en el i-ésimo punto de los datos. La diferencia entre e_i y ε_i se muestra en la siguiente figura.



MÉTODO PARA ENCONTRAR LOS COEFICIENTES DE REGRESIÓN

Mínimos cuadrados

Se encuentran a y b, estimaciones de α y β de modo que la suma de los cuadrados de los residuos sea mínima. A menudo la suma de cuadrados de los residuos se llama suma de cuadrados de los errores alrededor de la línea de regresión y se denota con SSE. Este procedimiento de minimización para estimar los parámetros se llama método de mínimos cuadrados.

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \bar{y} - b\bar{x}$$

LEOPOLDO VIVEROS ROSAS

Ejemplo 1

Uno de los problemas más desafiantes que enfrena el campo de control de contaminación del agua lo presenta la industria de curtido de pieles. Los deshechos de las curtidurías son químicamente complejos. Se caracterizan por los altos valores de demanda bioquímica de oxígeno, sólidos volátiles y otras medidas de contaminación. Se obtienen datos de 33 muestras de estos deshechos en un estudio que realizo la UPVM. Se registraron las lecturas de x, la reducción porcentual de sólidos totales y y la reducción porcentual en la demanda química de oxigeno para las 33 muestras. Hallar el modelo de regresión lineal mínimo que se ajusta a los datos. La tabla siguiente muestra las lecturas tomadas.

Reducción de sólidos	Demanda química de oxigeno	Reducción de sólidos	Demanda química de oxigeno
3	5	36	34
7	11	37	36
11	21	38	38
15	16	39	37
18	16	39	36
27	28	39	45
29	27	40	39
30	25	41	41
30	35	42	40
31	30	42	44
31	40	43	37
32	32	44	44
33	34	45	46
33	32	46	46
34	34	47	49
36	37	50	51
36	38		

Determinando las sumatorias que piden a y b se tienen las siguientes tablas.

xy	x^2
15	9
77	49
231	121
240	225
288	324
756	729
783	841
750	900
1050	900
930	961
1240	961
1024	1024
1122	1089
1056	1089
1156	1156
1332	1296
1368	1296

xy	x^2
1224	1296
1332	1369
1444	1444
1443	1521
1404	1521
1755	1521
1560	1600
1681	1681
1680	1764
1848	1764
1591	1849
1936	1936
2070	2025
2116	2116
2303	2209
2550	2500

$\sum_{i=1}^{n} x_i y_i$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i^2$
41355	**1104**	**1124**	**41086**

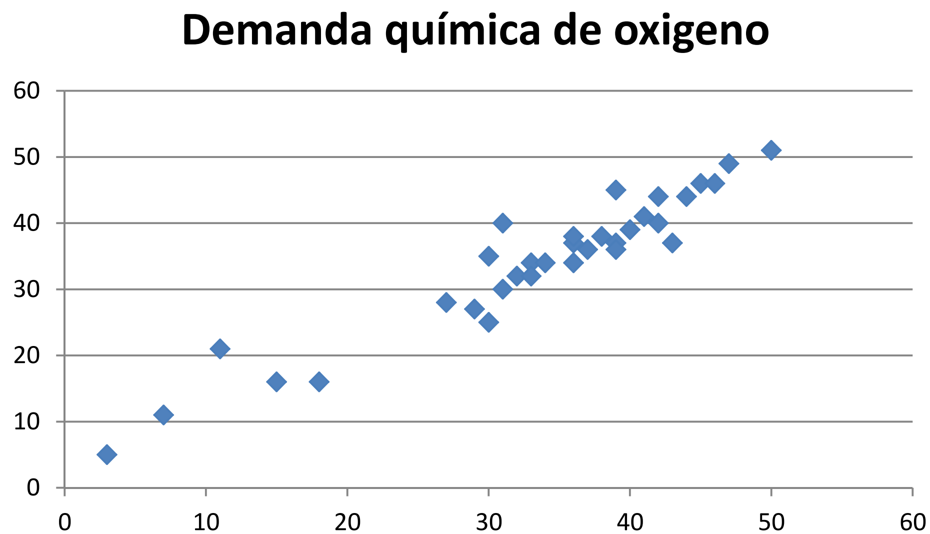
n=	33
b=	0.903643211
a=	3.829633198

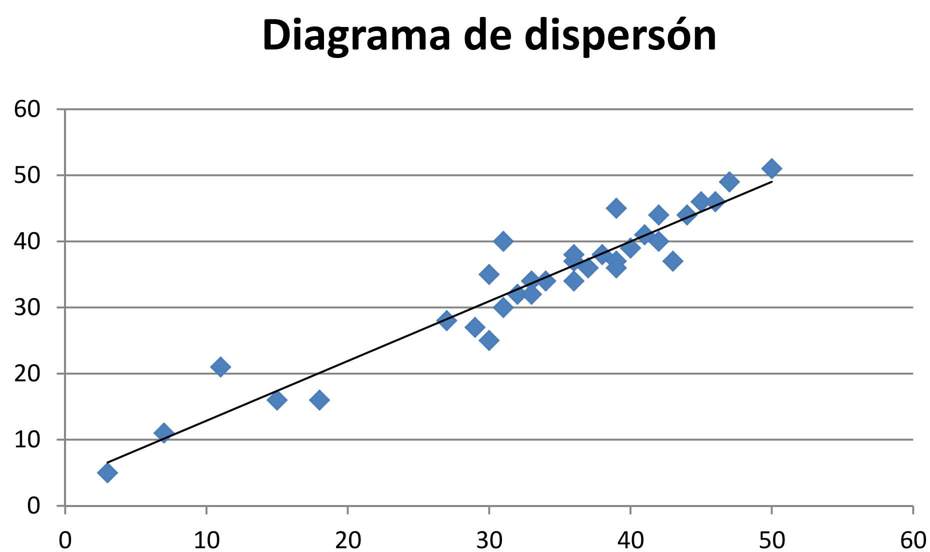
Por lo tanto el modelo queda como:

$$\hat{y} = 3.8296 + 0.9036x$$

LEOPOLDO VIVEROS ROSAS

El siguiente diagrama es el diagrama de dispersión de los datos.





LEOPOLDO VIVEROS ROSAS

Coeficiente de correlación
La medición ρ de la asociación lineal entre dos variables x y y se estima mediante el **coeficiente de correlación muestral** r, donde:

$$r = b\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Los valores de r deben ser entre -1 y 1, si r=1 es una relación positiva perfecta, y si r=-1 es una relación negativa perfecta.

Para nuestro ejemplo el coeficiente de correlación es : r= 0.955479, lo que indica una relación positiva muy fuerte entre las dos variables.

Inferencia de los coeficientes de regresión

Sean $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$ $\qquad S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$

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

Por tanto un intervalo de confianza para β es:

Un intervalo de confianza de $(1-\alpha)100\%$ para el parámetroβ en la línea de regresión $\mu_{Y/x} = \alpha + \beta\chi$ es:

$$b - \frac{t_{\alpha/2} s}{\sqrt{S_{xx}}} < \beta < b + \frac{t_{\alpha/2} s}{\sqrt{S_{xx}}}$$

Donde $t_{\alpha/2}$ es un valor de distribución t con n-2 grados de libertad
Y S es la desviación estándar

De nuestro ejemplo

$\sum_{i=1}^{n}(x_i-\bar{x})^2$	$\sum_{i=1}^{n}(y_i-\bar{y})^2$
927.479339	844.518825
699.842975	531.791552
504.206612	170.579431
340.570248	326.185491
238.842975	326.185491
41.661157	36.7309458
19.8429752	49.8521579
11.9338843	82.0945822
11.9338843	0.88246097
6.02479339	16.4885216
6.02479339	35.2764004
2.11570248	4.24609734
0.20661157	0.00367309
0.20661157	4.24609734
0.29752066	0.00367309
6.47933884	8.64003673
6.47933884	15.5188246

$\sum_{i=1}^{n}(x_i-\bar{x})^2$	$\sum_{i=1}^{n}(y_i-\bar{y})^2$
6.47933884	0.00367309
12.5702479	3.76124885
20.661157	15.5188246
30.7520661	8.64003673
30.7520661	3.76124885
30.7520661	119.67034
42.8429752	24.3976125
56.9338843	48.1551882
73.0247934	35.2764004
73.0247934	98.7915519
91.1157025	8.64003673
111.206612	98.7915519
133.297521	142.549128
157.38843	142.549128
183.479339	223.185491
273.752066	286.943067
4152.18182	3713.87879

$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$	
885.027548	-0.15426997
610.057851	6.87603306
293.269972	17.9063361
333.300275	16.3002755
279.118457	10.7548209
39.1184573	60.6639118
31.4517906	32.3305785
31.3002755	52.3608815
-3.24517906	50.7548209
9.96694215	84.9366391
-14.5785124	28.0578512
2.99724518	104.815427
0.02754821	137.84573
0.93663912	149.785124
-0.03305785	202.360882
7.48209366	280.269972
10.0275482	3752.09091

De nuestro ejemplo

$\sum_{i=1}^{n}(x_i-\bar{x})^2$	$\sum_{i=1}^{n}(y_i-\bar{y})^2$	$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$
4152.18182	3713.87879	3752.09091

Para determinar la desviación estándar se tiene:

$$s^2 = \frac{S_{yy} - bS_{xy}}{n-2} \qquad \therefore \qquad s = \sqrt{\frac{S_{yy} - bS_{xy}}{n-2}}$$

$$s^2 = \frac{S_{yy} - bS_{xy}}{n-2} = \frac{3713.88 - (0.903643)(3752.09)}{31}$$

$$= 10.4290$$

$$s = \sqrt{10.4290} = 3.2293$$

S_{xx}=4152.18 S_{xy}= 3752.09

$$s^2 = \frac{S_{yy} - bS_{xy}}{n-2} = \frac{3713.88 - (0.903643)(3752.09)}{31}$$

$$= 10.4290$$

$$s = \sqrt{10.4290} = 3.2293$$

Para determinar $t_{\alpha/2}$, se tiene un 95% de confianza por lo tanto α=0.5/2=0.025, por tanto $t_{0.025}$ =2.042 que se toma de las tablas de t student. Con 31 grados de libertad.

$$0.903643 - \frac{(2.042)(3.2295)}{\sqrt{4152.18}} < \beta < 0.903643 + \frac{(2.042)(3.2295)}{\sqrt{4152.18}}$$

$$0.801303 < \beta < 1.00598$$

Por lo tanto nuestro estimar esta dentro del intervalo y es un buen estimador

Intervalo de confianza para la intersección o α

Un intervalo de confianza de (1-α)100% para el parámetro α en la línea de regresión $\mu_{Y/x} = \alpha + \beta\chi$ es:

$$a - \frac{t_{\alpha/2} s \sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{nS_{xx}}} < \alpha < a + \frac{t_{\alpha/2} s \sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{nS_{xx}}}$$

Donde $t_{\alpha/2}$ es un valor de la distribución t con n-2 grados de libertad

De nuestro ejemplo

Encuentre el intervalo de confianza de (1-α)100% para el parámetro α en la línea de regresión $\mu_{Y/x} = \alpha + \beta\chi$ con base en los datos de la tabla .

Se tiene que S_{xx} = 4152.18 y s = 3.2295

Entonces tenemos que:

$$\sum_{i=1}^{n} x_i^2 = 41{,}086 \qquad y \qquad a = 3.8296$$

Con el uso de tabla de student, encontramos $t_{0.025} \cong 2.045$ para 31-2 grados de libertad. Por tanto, un intervalo de confianza de 95% para α es:

$$3.829633 - \frac{(2.045)(3.2295)\sqrt{41{,}086}}{\sqrt{(33)(4152.18}} < \alpha < 3.829633 + \frac{(2.045)(3.2295)\sqrt{41{,}086}}{\sqrt{(33)(4152.18}}$$

Concluimos que nuestra α es buena.

INGENIERÍA INDUSTRIAL
DISEÑO EXPERIMENTAL

ANOVA PARA MODELO DE REGRESIÓN SIMPLE

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0
Regresión	$SC_R = \hat{\beta} S_{xy}$	1	CM_R	CM_R/CM_E
Error o residual	$SC_E = S_{yy} - \hat{\beta} S_{xy}$	n-2	CM_E	
Total	S_{yy}	n-1		

LEOPOLDO VIVEROS ROSAS

INGENIERÍA INDUSTRIAL
DISEÑO EXPERIMENTAL

ANÁLISIS DE REGRESIÓN

Parámetro	Estimación	Error Estándar	Estadístico
intercepción	$\hat{\alpha} = \bar{y} - \beta\bar{x}$	$\sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}$	$\dfrac{\hat{\alpha}}{\sqrt{CM_E\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]}}$
pendiente	$\hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$	$\sqrt{\dfrac{CM_E}{S_{xx}}}$	$\dfrac{\hat{\beta}}{\sqrt{\dfrac{CM_E}{S_{xx}}}}$

LEOPOLDO VIVEROS ROSAS

Ejercicios 2

Es importante que los investigadores en el área de los productos forestales sean capaces de estudiar la relación entre la anatomía y las propiedades mecánicas de los árboles. De acuerdo con un estudio que llevo a cabo el departamento de silvicultura y productos forestales del IPN, se tomaron aleatoriamente 29 pinos loblolly para investigación que produjeron los datos de la siguiente tabla, sobre la gravedad específica en gramos/cm^3 y el módulo de ruptura en kilopascales, Ajuste a un modelo de regresión lineal y encuentre su correlación.

Gravedad específica	Módulo de ruptura	Gravedad específica	Módulo de ruptura
x (g/cm^3)	y (Kpa)	x (g/cm^3)	y (Kpa)
0.424	29,186	0.581	85,156
0.383	29,266	0.557	69,571
0.399	26,215	0.55	84,160
0.402	30,162	0.531	73,466
0.442	38,867	0.55	78,610
0.422	37,831	0.556	67,657
0.466	44,576	0.523	74,017
0.5	46,097	0.602	87,291
0.514	59,698	0.569	86,836
0.53	67,705	0.544	82,540
0.569	66,088	0.557	81,699
0.558	78,486	0.53	82,096
0.577	89,869	0.547	75,657
0.572	77,369	0.585	80,490
0.548	67,095		

LEOPOLDO VIVEROS ROSAS

Ejemplo 3

Un estudio de la cantidad de precipitación pluvial y la cantidad de contaminación eliminada del aire produce los siguientes datos.

Precipitación pluvial diaria,	Partículas eliminadas
x (0.01 cm)	y(mcg/m^3)
4.3	126
4.5	121
5.9	116
5.6	118
6.1	114
5.2	118
3.8	132
2.1	141
7.5	108

Regresión lineal múltiple

En la mayor parte de los problemas donde se aplica el análisis de regresión se necesita más de una variable independiente en el modelo de regresión. La complejidad de la mayor parte de los mecanismos científicos es tal que para ser capaces de predecir una respuesta importante se necesita un modelo de regresión múltiple. Cuando este modelo es lineal en los coeficientes se denomina **modelo de regresión lineal múltiple.** Para el caso de k variables independientes $x_1, x_2, \dots, x_k$ la media de $Y/ x_1, x_2, \dots, x_k$ esta dada por el modelo de regresión lineal múltiple.

$$\mu_{Y/ x1, x2, \dots, xk} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

Y la respuesta estimada se obtiene de la ecuación de regresión de la muestra

$$\hat{y} = b_0 + b_1 x_1 + \dots + b_k x_k$$

Estimación de los coeficientes

Por el método de mínimos cuadrados se obtienen las ecuaciones normales generales:

$$nb_0 + b_1\sum_{i=1}^{n} x_{1i} + b_2\sum_{i=1}^{n} x_{2i} + \ldots + b_k\sum_{i=1}^{n} x_{ki} = \sum_{i=1}^{n} y_i$$

$$b_0\sum_{i=1}^{n} x_{1i} + b_1\sum_{i=1}^{n} x_{1i}^2 + b_2\sum_{i=1}^{n} x_{1i}x_{2i} + \ldots + b_k\sum_{i=1}^{n} x_{1i}x_{ki} = \sum_{i=1}^{n} x_{1i}y_i$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$b_0\sum_{i=1}^{n} x_{ki} + b_1\sum_{i=1}^{n} x_{ki}x_{1i} + b_2\sum_{i=1}^{n} x_{ki}x_{2i} + \ldots + b_k\sum_{i=1}^{n} x_{ki}^2 = \sum_{i=1}^{n} x_{ki}y_i$$

Para dos variables

$$nb_0 + b_1 \sum_{i=1}^{2} x_{1i} + b_2 \sum_{i=1}^{2} x_{2i} = \sum_{i=1}^{2} y_i$$

$$b_0 \sum_{i=1}^{2} x_{1i} + b_1 \sum_{i=1}^{2} x_{1i} x_{1i} + b_2 \sum_{i=1}^{2} x_{1i} x_{2i} = \sum_{i=1}^{2} x_{1i} y_i$$

$$b_0 \sum_{i=1}^{2} x_{2i} + b_1 \sum_{i=1}^{2} x_{2i} x_{1i} + b_2 \sum_{i=1}^{2} x_{2i} x_{2i} = \sum_{i=1}^{2} x_{2i} y_i$$

Para tres variables

$$nb_0 + b_1 \sum_{i=1}^{3} x_{1i} + b_2 \sum_{i=1}^{3} x_{2i} + b_3 \sum_{i=1}^{3} x_{3i} = \sum_{i=1}^{3} y_i$$

$$b_0 \sum_{i=1}^{3} x_{1i} + b_1 \sum_{i=1}^{3} x_{1i} x_{1i} + b_2 \sum_{i=1}^{3} x_{1i} x_{2i} + b_3 \sum_{i=1}^{3} x_{1i} x_{2i} = \sum_{i=1}^{3} x_{1i} y_i$$

$$b_0 \sum_{i=1}^{3} x_{2i} + b_1 \sum_{i=1}^{3} x_{2i} x_{1i} + b_2 \sum_{i=1}^{3} x_{2i} x_{2i} + b_3 \sum_{i=1}^{3} x_{2i} x_{3i} = \sum_{i=1}^{3} x_{2i} y_i$$

$$b_0 \sum_{i=1}^{3} x_{3i} + b_1 \sum_{i=1}^{3} x_{3i} x_{1i} + b_2 \sum_{i=1}^{3} x_{3i} x_{2i} + b_3 \sum_{i=1}^{3} x_{3i} x_{3i} = \sum_{i=1}^{3} x_{3i} y_i$$

Ejemplo 4

Se realizo un estudio sobre un camión de reparto ligero a disel para ver si la humedad, temperatura del aire y presión barométrica influyen en la emisión de óxido nitroso (en ppm). Las mediciones de las emisiones se tomaron en diferentes momentos, con condiciones experimentales variantes. Los datos son los siguientes:

óxido nitroso	Humedad	Temperatura	Presión
y	x1	x2	x3
0.9	72.4	76.3	29.18
0.91	41.6	70.3	29.35
0.96	34.3	77.1	29.24
0.89	35.1	68	29.27
1	10.7	79	29.78
1.1	12.9	67.4	29.39
1.15	8.3	66.8	29.69
1.03	20.1	76.9	29.48
0.77	72.2	77.7	29.09
1.07	24	67.7	29.6
1.07	23.2	76.8	29.38
0.94	47.4	86.6	29.35
1.1	31.5	76.9	29.63
1.1	10.6	86.3	29.56
1.1	11.2	86	29.48
0.91	73.3	76.3	29.4
0.87	75.4	77.9	29.28
0.78	96.6	78.7	29.29
0.82	107.4	86.8	29.03
0.95	54.9	70.9	29.37

Ajuste el modelo de regresión lineal múltiple a los datos dados y después estime la cantidad de óxido nitroso para las condiciones donde la humedad es 50%, la temperatura 76°F y la presión barométrica 29.30.

Determinando los valores que se piden en las ecuaciones se tiene

x1*x1	x2*x2	x3*x3	x1*x2	x1*x3	x2*x3	x1*y1	x2*y2	x3*y3
5241.76	5821.69	851.4724	5524.12	2112.632	2226.434	65.16	68.67	26.262
1730.56	4942.09	861.4225	2924.48	1220.96	2063.305	37.856	63.973	26.7085
1176.49	5944.41	854.9776	2644.53	1002.932	2254.404	32.928	74.016	28.0704
1232.01	4624	856.7329	2386.8	1027.377	1990.36	31.239	60.52	26.0503
114.49	6241	886.8484	845.3	318.646	2352.62	10.7	79	29.78
166.41	4542.76	863.7721	869.46	379.131	1980.886	14.19	74.14	32.329
68.89	4462.24	881.4961	554.44	246.427	1983.292	9.545	76.82	34.1435
404.01	5913.61	869.0704	1545.69	592.548	2267.012	20.703	79.207	30.3644
5212.84	6037.29	846.2281	5609.94	2100.298	2260.293	55.594	59.829	22.3993
576	4583.29	876.16	1624.8	710.4	2003.92	25.68	72.439	31.672
538.24	5898.24	863.1844	1781.76	681.616	2256.384	24.824	82.176	31.4366
2246.76	7499.56	861.4225	4104.84	1391.19	2541.71	44.556	81.404	27.589
992.25	5913.61	877.9369	2422.35	933.345	2278.547	34.65	84.59	32.593
112.36	7447.69	873.7936	914.78	313.336	2551.028	11.66	94.93	32.516
125.44	7396	869.0704	963.2	330.176	2535.28	12.32	94.6	32.428
5372.89	5821.69	864.36	5592.79	2155.02	2243.22	66.703	69.433	26.754
5685.16	6068.41	857.3184	5873.66	2207.712	2280.912	65.598	67.773	25.4736
9331.56	6193.69	857.9041	7602.42	2829.414	2305.123	75.348	61.386	22.8462
11534.76	7534.24	842.7409	9322.32	3117.822	2519.804	88.068	71.176	23.8046
3014.01	5026.81	862.5969	3892.41	1612.413	2082.333	52.155	67.355	27.9015
54876.89	117912.32	17278.5086	67000.09	25283.395	44976.867	779.477	1483.437	571.1219

$$20b_0 \;+ b_1 863.1 + b_2 1530.4 + b_3 587.84 = 19.42$$

$$b_0 863.1 + b_1 54876.89 + b_2 67000.09 + b_3 25283.395 = 779.477$$

$$b_0 1539.4 + b_1 67000.09 + b_2 117912.32 + b_3 44976.867 = 1483.437$$

$$b_0 587.84 + b_1 25283.395 + b_2 44976.867 + b_3 17278.5086 = 571.1219$$

Encontrando la solución a este sistema se tiene:
b_0=-3.5077 b_1= -0.002625 b_2=0.000799, b_3=0.154155

Por lo tanto la ecuación de regresión queda como:

$$\hat{y} = -3.5077 - 0.002625x_1 + 0.000799x_2 + 0.15155x_3$$

Para 50% de humedad, una temperatura de 76 °F y una presión barométrica de 29.30, la cantidad estimada de óxido nitrosos es 0.9384

Ejemplo 5

En Applied Spectroscopy aparece un estudio sobre las propiedades espectrales de la reflectancia infraroja de un líquido viscoso que se utiliza en la industria de la electrónica como lubricante. El diseño experimental consiste en el efecto de frecuencia de banda x_1 y el espesor de la película x_2 sobre la densidad óptica y mediante el uso de un espectrómetro infrarojo de perkin-Elmer. Estimar la ecuación de regresión lineal

y	x_1	x_2
0.231	740	1.1
0.107	740	0.62
0.053	740	0.31
0.129	805	1.1
0.069	805	0.62
0.03	805	0.31
1.005	980	1.1
0.559	980	0.62
0.321	980	0.31
2.948	1235	1.1
1.633	1235	0.62
0.934	1235	0.31

Ejemplo 6

Se considera que la energía eléctrica que consume una planta química cada mes está relacionada con la temperatura ambiente promedio x_1, el numero de días al mes x_2, la pureza promedio del producto x_3 y las toneladas de producto fabricadas x_4. Se dispone de los datos históricos del año pasado y se presentan en la siguiente tabla.

y	x1	x2	x3	x4
240	25	24	91	100
236	31	211	90	95
290	45	24	88	110
274	60	25	87	88
301	65	25	91	94
316	72	26	94	99
300	80	25	87	97
296	84	25	86	96
267	75	24	88	110
276	60	25	91	105
288	50	25	90	100
261	38	23	89	98

Ajustar a la regresión lineal