FACULTAD DE INGENIERÍA
DIVISIÓN DE CIENCIAS BÁSICAS
COORDINACIÓN DE CIENCIAS APLICADAS
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA

SERIE TEMA 7
ANÁLISIS DE DATOS BIVARIADOS
PROBLEMAS CON RESOLUCIÓN

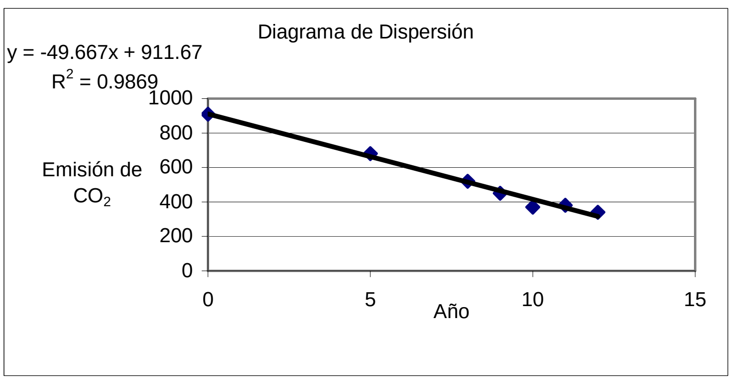
1. Los datos siguientes corresponden a la emisión de dióxido de carbono (CO_2) de calderas alimentadas por carbón (en unidades de 1000 ton) durante los años 1995 - 2007. La variable independiente (el año) se estandarizó para obtener la tabla siguiente:

Año	0	5	8	9	10	11	12
Emisión de CO_2	910	680	520	450	370	380	340

a) Estimar la recta de regresión.
b) ¿Existe una tendencia lineal significativa en las emisiones de CO_2 durante ese periodo? Justificar su respuesta.
c) ¿Sería aconsejable usar la recta de regresión estimada para calcular la emisión promedio de CO_2 de las calderas alimentadas con carbón en el año 2030? Justificar la respuesta.

Resolución

a) El diagrama de dispersión es:



La recta de regresión está dada por: $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$

donde: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ y $\hat{\beta}_1 = \dfrac{SS_{xy}}{SS_{xx}}$

se sabe que:

$$\bar{x} = \frac{1}{n} \sum_{n=1}^{7} x_i \quad , \quad \bar{y} = \frac{1}{n} \sum_{n=1}^{7} y_i \quad ,$$

$$SS_{xy} = \sum_{n=1}^{7} x_i \, y_i - \frac{\sum_{n=1}^{7} x_i \sum_{n=1}^{7} y_i}{n}$$

y $\quad SS_{xx} = \sum_{n=1}^{7} x_i^2 - \frac{\left(\sum_{n=1}^{7} x_i\right)^2}{n}$

sustituyendo en cada caso:

$\bar{x} = \frac{55}{7} = 7.857$

$\bar{y} = \frac{3650}{7} = 521.428$

$SS_{xy} = 23570 - \frac{200750}{7} = -\frac{35760}{7} = -5108.571$

$SS_{xx} = 535 - \frac{3025}{7} = \frac{720}{7} = 102.857$

$\hat{\beta}_1 = \frac{-\frac{35760}{7}}{\frac{720}{7}} = -\frac{35760}{720} = -\frac{149}{3} = -49.667$

$\hat{\beta}_0 = \frac{3650}{7} + \frac{149}{3}\left(\frac{55}{7}\right) = \frac{2735}{3} = 911.667$

por lo tanto el ajuste está dado por:

$\hat{y} = 911.667 - 49.667\,x$

b) Se requiere $SS_{yy} = \sum_{n=1}^{7} y_i^2 - \frac{\left(\sum_{n=1}^{7} y_i\right)^2}{n}$

sustituyendo:

$$SS_{yy} = 2160300 - \frac{(3650)^2}{7} = 2160300 - \frac{13322500}{7} = \frac{1799600}{7} = 257085.714$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-5108.57}{\sqrt{102.857(257085.714)}} = -\frac{5108.57}{5142.282} - 0.993$$

$r^2 = (-0.993)^2 = 0.987$

La tendencia lineal es muy buena.

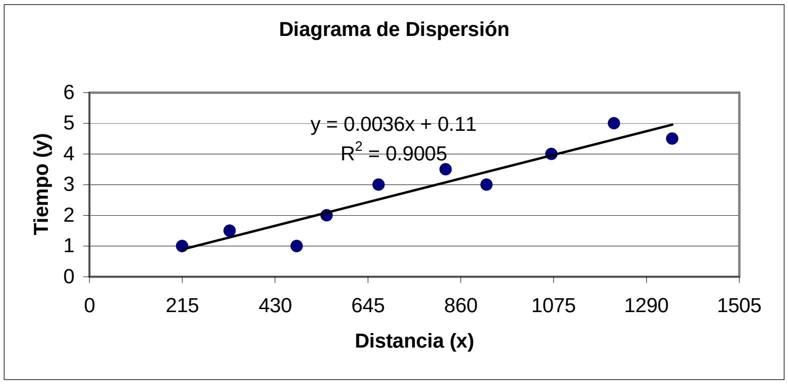
c) No, porque el año 2030 está muy alejado de la información que se tiene.

2. Supóngase que un ingeniero toma una muestra aleatoria de 10 embarques recientemente enviados por camión de una compañía y registra la distancia en kilómetros y el tiempo de entrega, al mediodía más cercano, y a partir del momento en que el embarque estuvo listo para su transportación.

Distancia (x), [Km]	825	215	1070	550	480	920	1350	325	670	1215
Tiempo de entrega (y), [días]	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

a) Construir la gráfica de dispersión.
b) Estimar la recta de regresión.
c) Calcular el coeficiente de determinación e interpretar el resultado.

Resolución
a) El diagrama de dispersión es:



b) La recta de regresión está dada por: $\hat{y} = \hat{\beta}_1 x + \hat{\beta}_0$

donde:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

y

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

Con $n = 10$, se sabe que:

$$\bar{x} = \frac{1}{n}\sum_{n=1}^{10} x_i \quad , \quad \bar{y} = \frac{1}{n}\sum_{n=1}^{10} y_i \quad ,$$

$$SS_{xy} = \sum_{n=1}^{10} x_i y_i - \frac{\sum_{n=1}^{10} x_i \sum_{n=1}^{10} y_i}{n}$$

y

$$SS_{xx} = \sum_{n=1}^{10} x_i^2 - \frac{\left(\sum_{n=1}^{10} x_i\right)^2}{n}$$

sustituyendo en cada caso:

$$\bar{x} = \frac{7620}{10} = 762$$

$$\bar{y} = \frac{28.5}{10} = 2.85$$

$$SS_{xy} = 26370 - \frac{(7620)(28.5)}{10} = 4653$$

$$SS_{\text{xx}} = 7104300 - \frac{(7620)^2}{10} = 1297860$$

sustituyendo:

$$\hat{\beta}_1 = \frac{4653}{1297860} = 0.0036$$

$$\hat{\beta}_0 = 2.85\text{-}(0.0036)\ (762) = 0.11$$

por lo tanto el ajuste a una recta está dado por:

$$\hat{y} = 0.0036x + 0.11$$

c) Se requiere $SS_{yy} = \sum_{n=1}^{10} y_i^2 - \frac{\left(\sum_{n=1}^{10} y_i\right)^2}{n}$

sustituyendo:

$$SS_{\text{yy}} = 99.75 - \frac{(28.5)^2}{10} = 18.525$$

se sabe que:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{4653}{\sqrt{(1297860)(18.525)}} = 0.9489$$

entonces:

$$r^2 = (0.9489)^2 = 0.9004$$

La tendencia lineal es muy buena.

3. En la producción de herramientas de acero, se ha considerado ilustrar la relación entre la deformación (x) y la dureza Brinell (y).

a) Obtener la ecuación de la recta de regresión.
b) Calcular el coeficiente de determinación.
c) La dureza cuando la deformación es de 25 [mm]

$x\ [mm]$	6	9	11	22	26	28	33	35
$y\ \left[\frac{kg}{mm^2}\right]$	68	67	65	44	40	37	34	32

Usar los cálculos siguientes

	x	y	x^2	y^2	xy
Suma:	170	387	4496	20423	7012

Resolución

a) El ajuste de los datos a un modelo lineal de regresión por el criterio de mínimos cuadrados está dado por

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

donde

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

realizando los productos y las sumas, se tiene

	x	y	x^2	y^2	xy
	6	68	36	4624	408
	9	67	81	4489	603
	11	65	121	4225	715
	22	44	484	1936	968
	26	40	676	1600	1040
	28	37	784	1369	1036
	33	34	1089	1156	1122
	35	32	1225	1024	1120
Suma:	**170**	**387**	**4496**	**20423**	**7012**

de donde

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 4496 - \frac{(170)^2}{8} = 883.5$$

$$SS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 7012 - \frac{(170)(387)}{8} = -1211.75$$

sustituyendo

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{-1211.75}{883.5} = -1.372$$

para calcular los promedios, se tiene

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(170) = 21.25$$

y

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{8}(387) = 48.375$$

sustituyendo

$$\hat{\beta}_0 = 48.375 - (-1.372)(21.25) = 77.53$$

por lo tanto el modelo lineal de regresión es

$$\hat{y} = 77.53 - 1.372\,x$$

b) El coeficiente de correlación está definido por

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}$$

calculando

$$SS_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 20423 - \frac{(387)^2}{8} = 1701.875$$

por lo que el coeficiente de correlación es

$$r = \frac{-1211.75}{\sqrt{(883.5)(1701.875)}} = -0.988$$

El coeficiente de determinación está definido por

$R^2 = (r)^2 = (-0.988)^2 = 0.976$

Se concluye que las variables tienen una buena relación lineal, puesto que: $R^2 = 0.976$

c) Si la deformación es de 25 [mm], entonces

$\hat{y} = 77.53 - 1.372(25) = 43.23$

4. Se realizó un estudio para determinar los efectos de no dormir en la capacidad de las personas para resolver problemas sencillos. La cantidad variaba de 8, 12, 16, 20 a 24 horas sin dormir. Cinco personas participaron en el estudio. Se dio a cada persona, después de un periodo específico sin dormir, un conjunto de problemas sencillos de sumar y se registro el número de errores. Se obtuvieron los siguientes resultados

número de errores, (y)	8	10	14	12	16
número de horas sin *dormir*, (x)	8	12	16	20	24

a) Determinar la recta apropiada de mínimos cuadrados para estos datos.

b) Trazar el diagrama de dispersión y la recta del inciso (a).

Resolución

a) El ajuste de los datos a un modelo lineal de regresión por el criterio de mínimos cuadrados está dado por

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

donde

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

realizando los productos y las sumas, se tiene

	x	y	x^2	y^2	xy
	8	8	64	64	64
	12	10	144	100	120
	16	14	256	196	224
	20	12	400	144	240
	24	16	576	256	384
Suma:	80	60	1440	760	1032

de donde

$$SS_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 1440 - \frac{(80)^2}{5} = 160$$

$$SS_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 1032 - \frac{(80)(60)}{5} = 72$$

sustituyendo

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{72}{160} = 0.45$$

para calcular los promedios, se tiene

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(80) = 16$$

y

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i=\frac{1}{5}(60)=12$$

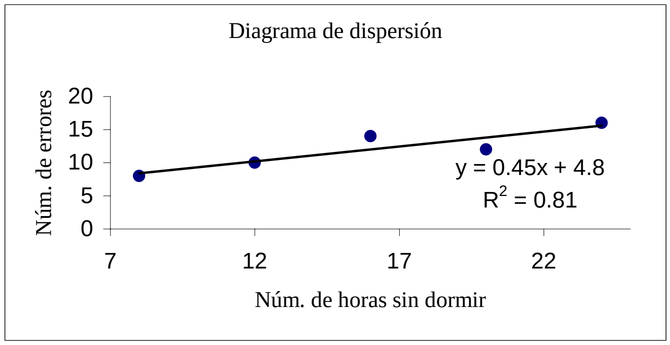
sustituyendo

$$\hat{\beta}_0=12-(0.45)(16)=4.8$$

por lo tanto el modelo lineal de regresión es

$$\hat{y}=0.45x+4.8$$

b) El diagrama de dispersión y la recta del inciso (a) son



5. La siguiente tabla muestra datos sobre el desgaste de acero dulce y y la viscosidad del aceite x.

y $[10^{-4}\ mm^3]$	240	181	193	155	172	110	113	75	94
x	1.6	9.4	15.5	20	22	35.5	43	40.5	33

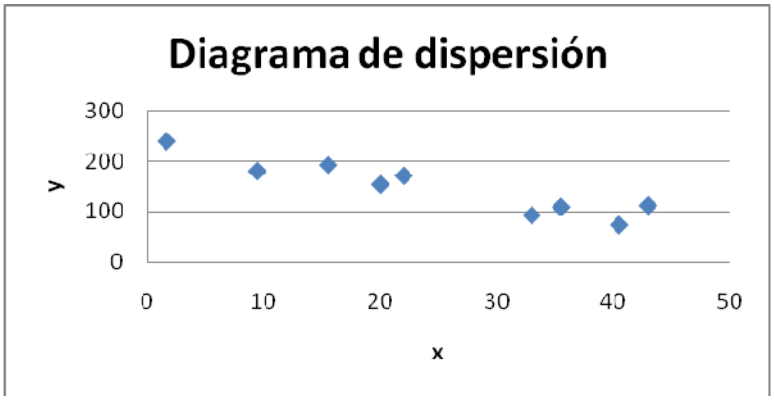
Las sumas relevantes, se muestran a continuación:

Sumas	
y	1333
x	220.5
xy	26864.4
x^2	7053.67
y^2	220549

a) Trazar la gráfica de la dispersión de los datos. ¿Parece conveniente el uso de un modelo de regresión lineal?
b) Ajustar un modelo de regresión lineal simple.
c) Determinar el valor que se espera del desgaste cuando la viscosidad es 30
d) ¿Puede considerarse válido el modelo? Justificar su respuesta.

Resolución

a)



b) El modelo es:

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{9} x_i y_i - \frac{\sum_{i=1}^{9} x_i \sum_{i=1}^{9} y_i}{9}}{\sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{9}} = \frac{26864.4 - \frac{(220.5)(1333)}{9}}{7053.67 - \frac{(220.5)^2}{9}} \approx -3.5086$$

y

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \frac{1333}{9} - (-3.5086)\left(\frac{220.5}{9}\right) \approx 234.0718$$

El modelo queda:

$$\hat{y} = 234.0718 - 3.5086x$$

c) Para el valor que se espera del desgaste cuando la viscosidad es 30, sustituyendo en el modelo:

$$\hat{y}(30) = 234.0718 - 3.5086(30) \approx 128.8138$$

d) Para determinar si el modelo es válido debe obtenerse el coeficiente de determinación. El coeficiente de correlación, está definido por:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}$$

$$SS_{xx} = \sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{9} = 7053.67 - \frac{(220.5)^2}{9} = 1651.42$$

$$SS_{yy} = \sum_{i=1}^{9} y_i^2 - \frac{\left(\sum_{i=1}^{9} y_i\right)^2}{9} = 220549 - \frac{(1333)^2}{9} \approx 23116.8889$$

$$SS_{xy}=\sum_{i=1}^{9}x_iy_i-\frac{\sum_{i=1}^{9}x_i\sum_{i=1}^{9}y_i}{9}=26864.4-\frac{(220.5)(1333)}{9}=-5794.1$$

sustituyendo:

$$r=\frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}}=\frac{-5794.1}{\sqrt{(1651.42)(23116.8889)}}\approx -0.9378$$

Entonces el coeficiente de determinación será:

$$r^2=R^2=\frac{SS_{xy}^2}{SS_{xx}SS_{yy}}=\frac{(-5794.1)^2}{(1651.42)(23116.8889)}\approx 0.8794$$

El ajuste es regular y puede considerarse válido el modelo dependiendo del error que se esté dispuesto a cometer.

6. La siguiente tabla muestra la relación que se presenta entre la pureza del oxígeno producido en un proceso de destilación química, contra el porcentaje de hidrocarburos que están presentes en el condensador principal de la unidad de destilación.

Nivel de hidrocarburos	x (%)	0.99	1.02	1.05	1.29	1.46	1.36	0.87	1.23	1.55	1.4
Pureza	y (%)	90.01	89.05	91.43	93.74	96.73	94.45	87.59	91.77	99.42	93.65
	x (%)	1.19	1.15	0.98	1.01	1.11	1.2	1.26	1.32	1.43	0.95
	y (%)	93.54	92.52	90.56	89.54	89.85	90.39	93.25	93.41	94.98	87.33

a) Ajustar un modelo de regresión lineal simple.

b) ¿Puede considerarse válido el modelo? Justificar su respuesta.

Se sabe que:

Sumas:

x	y	x^2	y^2	xy
23.82	1843.21	29.0692	170044.5321	2205.5136

Resolución

a) El modelo es:

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$$

$$\hat{\beta}_1=\frac{\sum_{i=1}^{20}x_iy_i-\frac{\sum_{i=1}^{20}x_i\sum_{i=1}^{20}y_i}{20}}{\sum_{i=1}^{20}x_i^2-\frac{\left(\sum_{i=1}^{20}x_i\right)^2}{20}}=\frac{2205.5136-\frac{(23.82)(1843.21)}{20}}{29.0692-\frac{(23.82)^2}{20}}\approx 14.6523$$

y

$$\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}$$

$$\hat{\beta}_0=\frac{1843.21}{20}-(14.6523)\left(\frac{23.82}{20}\right)\approx 74.7096$$

El modelo queda:

$$\hat{y}=74.7096+14.6523x$$

b) Para determinar si el modelo es válido debe obtenerse el coeficiente de determinación. El coeficiente de correlación, está definido por:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}}$$

$$SS_{xx} = \sum_{i=1}^{20} x_i^2 - \frac{\left(\sum_{i=1}^{20} x_i\right)^2}{20} = 29.0692 - \frac{(23.82)^2}{20} \approx 0.6996$$

$$SS_{yy} = \sum_{i=1}^{20} y_i^2 - \frac{\left(\sum_{i=1}^{20} y_i\right)^2}{20} = 170044.5321 - \frac{(1843.21)^2}{20} \approx 173.3769$$

$$SS_{xy} = \sum_{i=1}^{20} x_i y_i - \frac{\sum_{i=1}^{20} x_i \sum_{i=1}^{20} y_i}{20} = 2205.5136 - \frac{(23.82)(1843.21)}{20} \approx 10.2505$$

sustituyendo:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}} = \frac{10.2505}{\sqrt{(0.6996)(173.3769)}} \approx 0.9307$$

Entonces el coeficiente de determinación será:

$$r^2 = R^2 = \frac{SS_{xy}^2}{SS_{xx}SS_{yy}} = \frac{(10.2505)^2}{(0.6996)(173.3769)} \approx 0.8663$$

El ajuste es regular y puede considerarse válido el modelo dependiendo del error que se esté dispuesto a cometer.

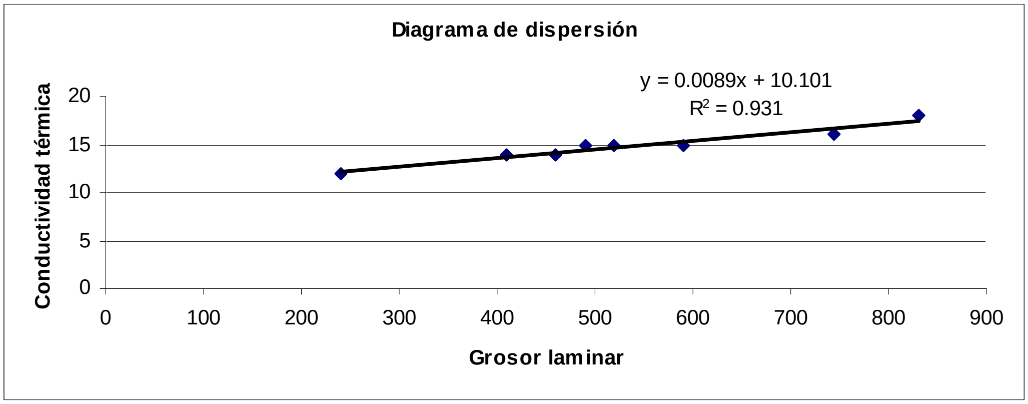
7. Una gráfica que aparece en el artículo "Thermal conductivity of polyethilenc: The effects of cristal size, density and orientation on the thermal conductivity" sugiere que el valor esperado de conductividad térmica y es una función lineal de $\frac{10^4}{x}$ donde x es el grosor laminar [en ångström].

x	240	410	460	490	520	590	745	830
y	12	14	14	15	15	15	16	18

a) Trazar el diagrama de dispersión.
b) Estimar los parámetros de la función de regresión y su función de regresión.
c) Pronosticar el valor de conductividad térmica cuando el grosor laminar es de 500 [Å]

Resolución

a) El diagrama de dispersión es



b) Los parámetros y el modelo, son

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{8} x_i y_i - \frac{\sum_{i=1}^{8} x_i \sum_{i=1}^{8} y_i}{8}}{\sum_{i=1}^{8} x_i^2 - \frac{\left(\sum_{i=1}^{8} x_i\right)^2}{8}} = \frac{65920 - \frac{(4285)(119)}{8}}{2539825 - \frac{(4285)^2}{8}} \approx 0.0089$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = \frac{119}{8} - (0.0089)\left(\frac{4285}{8}\right) \approx 10.1079$$

El modelo está dado por
$\hat{y} = 10.1079 + 0.0089x$

c) Para obtener la estimación del valor de conductividad térmica cuando el grosor laminar es de 500 [Å], es
$\hat{y} = 10.1079 + 0.0089(500) \approx 14.5579$

8. La Procuraduría del Consumidor evalúa anualmente distintas marcas de cigarros de acuerdo con el contenido de alquitrán, nicotina y monóxido de carbono (CO). La asociación de médicos considera peligrosas cada una de estas sustancias en la salud del fumador. Estudios anteriores han demostrado que un aumento en el contenido de alquitrán y nicotina de un cigarro está acompañado de un incremento en el monóxido de carbono emitido en el humo del cigarro. La tabla siguiente muestra los valores para seis marcas de cigarros.

Marca	**Alquitrán [mg]**	**CO [mg]**
Benson&Hedges	16,0	16,6
Camel Lights	8,0	10,2
Marlboro	15,1	14,4
Raleigh	15,8	17,5
Montana	17,0	18,5
Viceroy Light	8,6	10,6

a) Elaborar un diagrama de dispersión entre el contenido de alquitrán (x), y CO (y)

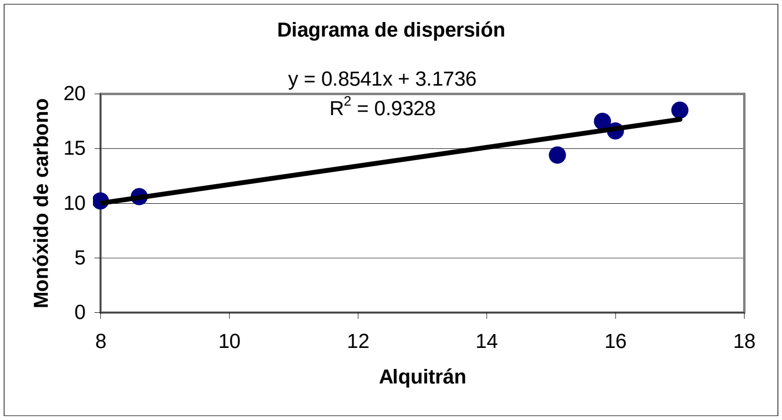
b) Obtener la recta de regresión y trazarla en el diagrama de dispersión del inciso (a)

c) ¿Corroboran los resultados el hecho de que un aumento de alquitrán conlleva un aumento de monóxido de carbono?

d) ¿Considera que el modelo proporcionado es bueno? ¿Qué porcentaje del monóxido de carbono emitido en el humo del cigarro queda explicado por el modelo?

Resolución

a) El diagrama de dispersión es:



b) Los parámetros y el modelo, son:

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{6} x_i y_i - \frac{\sum_{i=1}^{6} x_i \sum_{i=1}^{6} y_i}{6}}{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{6}} = \frac{1246.80 - \frac{(80.50)(87.80)}{6}}{1160.61 - \frac{(80.50)^2}{6}} = \frac{68.817}{80.568} \approx 0.8541$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} = 14.633 - (0.854)(13.417) \approx 3.1741$$

Por lo tanto el modelo está dado por:
$\hat{y} = 0.8541x + 3.1741$

c) De acuerdo a la gráfica, sí se corrobora el hecho de que un aumento de alquitrán conlleva un aumento de monóxido de carbono.

d) Sí es bueno el modelo que se obtuvo, ya que, el porcentaje del monóxido de carbono emitido en el humo del cigarro queda explicado por el modelo de manera muy aceptable como se observa en lo que sigue.
Para determinar si el modelo es válido debe obtenerse el coeficiente de determinación.
El coeficiente de correlación, está definido por:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}}$$

$$SS_{xx} = \sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{6} = 1160.61 - \frac{(80.50)^2}{6} = 80.568$$

$$SS_{yy}=\sum_{i=1}^{6}y_i^2-\frac{\left(\sum_{i=1}^{6}y_i\right)^2}{6}=1347.82-\frac{(87.80)^2}{6}=63.013$$

$$SS_{xy}=\sum_{i=1}^{6}x_iy_i-\frac{\sum_{i=1}^{6}x_i\sum_{i=1}^{6}y_i}{6}=1246.80-\frac{(80.50)(87.80)}{6}=68.817$$

sustituyendo:

$$r=\frac{SS_{xy}}{\sqrt{SS_{xx}\,SS_{yy}}}=\frac{68.817}{\sqrt{(80.568)(63.013)}}\approx 0.9658$$

Entonces el coeficiente de determinación es:

$$r^2=R^2=\frac{SS_{xy}^2}{SS_{xx}SS_{yy}}=\frac{(68.817)^2}{(80.568)(63.013)}\approx 0.9328$$

El ajuste es bueno puede considerarse válido el modelo, como se mencionó anteriormente.

9. Se llevó a cabo una investigación para estudiar la relación entre la velocidad [en pies por segundo] y el ritmo de zancadas [en número de pasos por segundo] de entre 11 corredores de maratón, se tienen los siguientes datos:

$$\sum_{n=1}^{11}(Velocidad)=205.4 \qquad \sum_{n=1}^{11}(Velocidad)^2=3880.08$$

$$\sum_{n=1}^{11}(Ritmo)=35.16 \qquad \sum_{n=1}^{11}(Ritmo)^2=112.681$$

$$\sum_{n=1}^{11}(Velocidad)(Ritmo)=660.13$$

Obtener la ecuación de la recta de regresión que se utilizaría para pronosticar el ritmo de zancadas a partir de la velocidad.

Resolución

Del enunciado se tiene $n=11$ y se sabe que la recta de regresión es

$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x$

donde

$$\hat{\beta}_1=\frac{\sum_{i=1}^{11}x_iy_i-\frac{\sum_{i=1}^{11}x_i\sum_{i=1}^{11}y_i}{11}}{\sum_{i=1}^{11}x_i^2-\frac{\left(\sum_{i=1}^{11}x_i\right)^2}{11}}=\frac{660.13-\frac{(205.4)(35.16)}{11}}{3880.08-\frac{(205.4)^2}{11}}=\frac{3.5969}{44.7018}\approx 0.0805$$

$$\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}$$

$$\overline{x}=\frac{\sum_{n=1}^{11}(Velocidad)}{n}=\frac{205.4}{11}=18.6727$$

y

$$\overline{y}=\frac{\sum_{n=1}^{11}(Ritmo)}{n}=\frac{35.16}{11}=3.1964$$

entonces

$$\hat{\beta}_0=3.1964-(0.0805)(18.6727)\approx 1.6932$$

por lo tanto el modelo está dado por

$$\hat{y}=0.0805x+1.6932$$