UNIVERSIDAD NACIONAL DEL LITORAL

FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

ESTADÍSTICA

RESPUESTAS DE GUÍA DE PRÁCTICA

UNIDAD 7 – REGRESIÓN Y CORRELACIÓN

Responsable de cátedra: Prof. Juan Pablo Taulamet

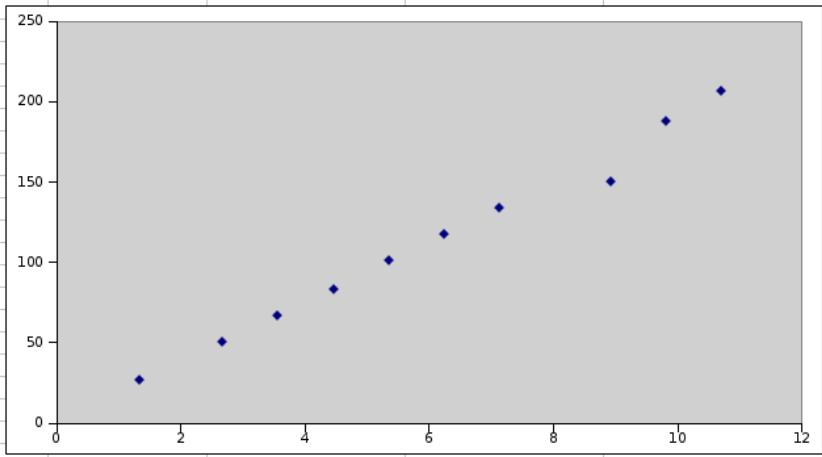
Equipo de cátedra: ***Auxiliares:*** Lic. María José Llop (JTP) - Ing. Ana Lisa Eusebi (JTP) - Prof. Fátima Bolatti (JTP) - Ing. Franco Nardi (Ay. 1°) ***Ayudantes:*** AIA Cristian Bottazzi - Téc. Eliana García

Carreras: Ingenierías en: Recursos Hídricos - Ambiental - Agrimensura

AÑO ACADÉMICO 2024

Ejercicio 1

A	B	C	D	E	F	G
x	y					
1,33	27					
2,68	50					
3,57	67					
4,46	83					
5,35	101					
6,24	117					
7,14	134					
8,93	150					
9,82	188					
10,7	206					
SALIDA RESUMEN						
Estadísticas de regresión						
R múltiple	0,9286816639417					
R^2	0,8624496329414					
Error estándar	9,7156543471832					
R^2 ajustado	0,8452558370591					
Observaciones	10					
Análisis de varianza						
	df	*SS*	*MS*	*F*		
Regresión	1	4734,84848484848	4734,8484848485	50,1605136436597	0,00010378187886	
Residual	8	755,151515151515	94,393939393939			
Total	9	5490				
	Coeficientes	*Error estándar*	*Estadísticas-t*	*Valor-P*	*0,95*	*0,95*
Interceptar	19,030303030303	10,6160558962146	1,7925963480551	0,11079931105051	−5,45036576592291	43,510971826536
Columna 1	0,7575757575758	0,10696583390741	7,0824087458759	0,00010378187886	0,51091210225967	1,00423941289185



a) De la observación del gráfico de dispersión parece adecuado proponer un modelo lineal para establecer una relación entre las variables velocidad y esfuerzo.

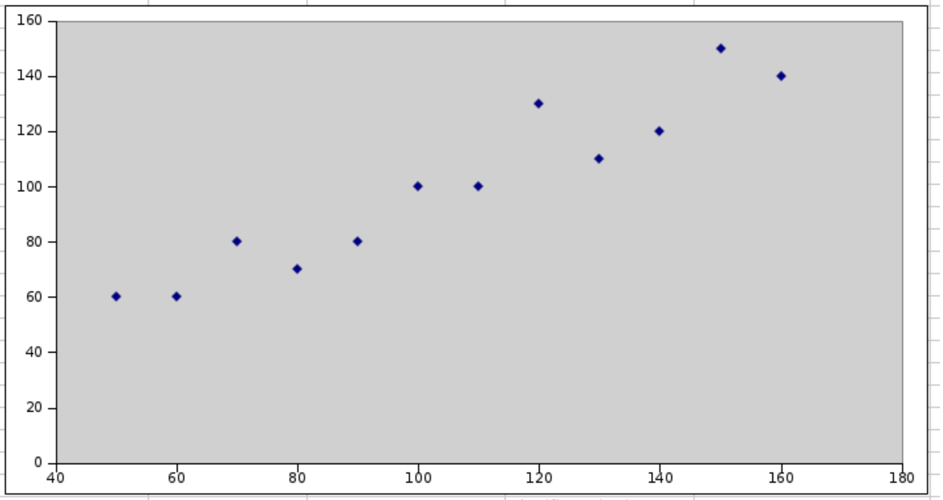
Se observa que el coeficiente de determinación r^2 es 0.989, es decir que el modelo propuesto sirve para explicar el 98% de la variabilidad de Y. A su vez por tratarse de un modelo lineal, cobra especial sentido analizar el valor del coeficiente de correlación $r = 0,928$ que implica un alto grado de asociación lineal entre las variables. Considerando que el coreficiente de correlación es positivo, la relación entre X e Y es directamente proporcional.

Además, la recta que resulta de la estimación por mínimos cuadrados es $\hat{y} = 0.698 + 18.532x$.

b) El valor estimado por mínimos cuadrados del coeficiente de regresión (pendiente de la recta b) es de 18.532. En la tabla además se presenta un intervalo del 95% para ese coeficiente. Como 20 es un valor perteneciente al intervalo calculado, 20 es un valor posible para la pendiente β con esa confianza.

c) Cuando x = 14, el valor esperado de y resulta 0.698+18.532 14 = 260.151

Ejercicio 2



x	y					
50	60					
60	60					
70	80					
80	70					
90	80					
100	100					
110	100					
120	130					
130	110					
140	120					
150	150					
160	140					
SALIDA RESUMEN						
Estadísticas de regresión						
R múltiple	0,9512031960481					
R^2	0,9047875201721					
Error estándar	9,9509285506921					
R^2 ajustado	0,8952662721893					
Observaciones	12					
Análisis de varianza						
	df	*SS*	*MS*	*F*	*Significancia de F*	
Regresión	1	9409,79020979021	9409,7902097902	95,0282485875707	2,00703661031E−06	
Residual	10	990,20979020979	99,020979020979			
Total	11	10400				
	Coeficientes	*Error estándar*	*Estadísticas-t*	*Valor-P*	*0,95*	*0,95*
Interceptar	14,825174825175	9,19754563531635	1,6118620567915	0,13806834514435	−5,66823394779323	35,3185835981429
Columna 1	0,8111888111888	0,08321384492668	9,7482433590658	2,007036610315E−06	0,62577681028449	0,99660081209313
t_(1-alfa/2), 10=	2,2281388519866	con alfa=0.05	promedio de x	105	promedio x - x^ast	92,5
SXX=	14300					
centro=	24,965034965035					
IC	6,6590228855954	43,2710470444746				

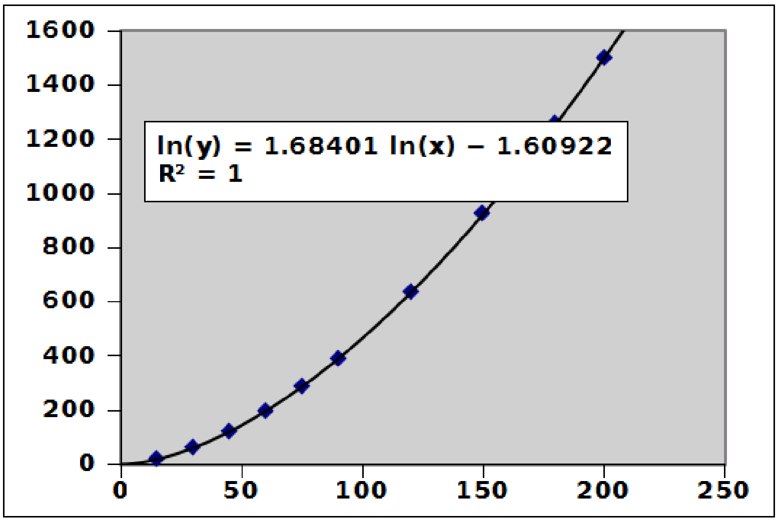
a) La recta obtenida por el método de mínimos cuadrados es $\hat{y} = 14.82 + 0.81x$. El valor obtenido para el coeficiente de determinación $r^2 = 0.9047$, pero por tratarse de un modelo lineal, le daremos más importancia al valor del coeficiente de correlación $r = 0.95$ que implica un fuerte grado de relación lineal proporcional, por lo que el ajuste obtenido puede servir para ser usado como herramienta de pronóstico.

b) De la tabla resulta que la varianza de la regresión estimada es 99.02.

c) La estimación puntual de la pendiente b es 0.81 y el IC para β con $1-\alpha = 95\%$ resulta $(0.63, 0.99)$.
d) El valor de la esperanza estimada de y cuando x=12.5, es 24.96. Un intervalo del 95% de confianza para el parámetro en estudio resulta (6.66, 43.27). Observación: el valor de x para el cual se desea estimar no pertenece al rango de valores de x con el cual se realizó el ajuste. Luego, los valores estimados arriba son válidos si el modelo es válido para valores de x tan pequeños como 12.5.

Ejercicio 3

A partir del uso de un software puede encontrarse una relación logarítmica en el dispersiograma:

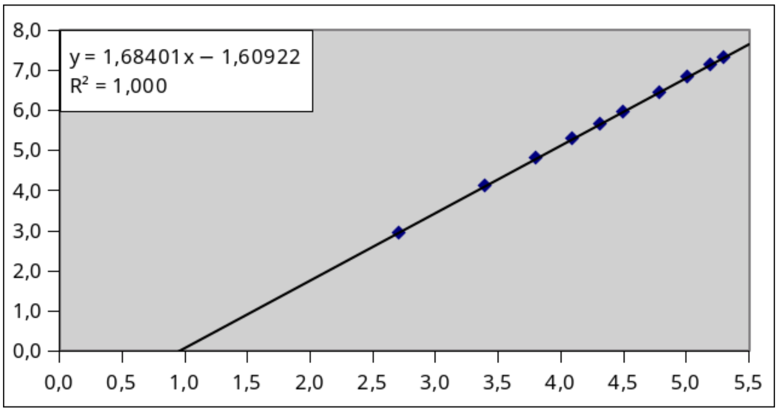


El coeficiente de determinación es igual a 1: todos los puntos están situados exactamente sobre la curva.

A su vez, si se estudia la relación lineal entre $ln(Q)$ y $ln(h)$, se pueden construir la siguiente tabla:

$ln(h)$	$ln(Q)$
2.7	3,0
3.4	4,1
3.8	4,8
4.1	5,3
4.3	5,7
4.5	6,0
4.8	6,5
5.0	6,8
5.2	7,1
5.3	7,3

A partir de lo anterior se puede obtener el siguiente dispersiograma:



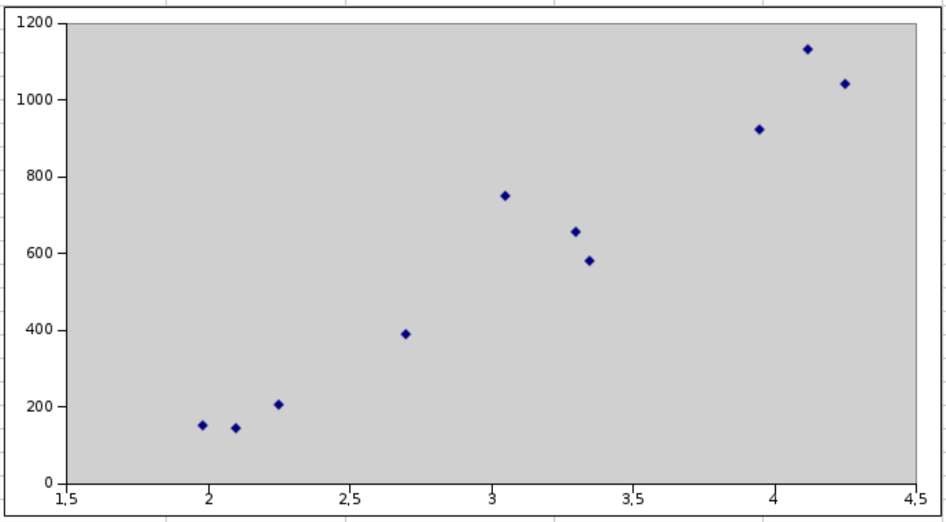
En ese caso la relación lineal es perfecta ya que posee un coeficiente de correlación igual a 1. A continuación se presenta el análisis de regresión:

SALIDA RESUMEN		Variable de respuesta	Columna 2			
Estadísticas de regresión						
R múltiple	0,99999999931363269					
R^2	0,99999999862726539					
Error están	5,4942110338676E−05					
R^2 ajusta	0,99999999845567356					
Observacic	10					
Análisis de varianza						
	df	SS	MS	F	Significancia de F	
Regresión	1	17,59195378327761417	17,59195378327761417	5827783397,7548444374	9,70966872611827E−37	
Residual	8	2,414908390773835E−08	3,01863548846729E−09			
Total	9	17,591953807426698078				
	Coeficientes	Error estándar	Estadísticas-t	Valor-P	0,95	0,95
Interceptaı	−1,6092203213253393	9,668694044027836E−05	−16643,6161284814192	1,9021136090551760E−31	−1,60944328180981481	−1,6089973608408637646
Columna 1	1,68400721230698183	2,205932637295711E−05	76339,9200789482843	9,7096687267489471E−37	1,683956343409145974	1,6840580812048176889

b) El modelo de regresión permite estimar en forma puntual que si $\hat{Q} = exp(1.68401 * ln(400) - 1.60922) = 4819.70 m^3/s$ cuando $h = 400 cm$.

Ejercicio 4

x	y
1,98	150
2,1	144
2,25	205
2,7	388
3,05	750
3,3	653
3,35	580
3,95	920
4,12	1130
4,25	1040



SALIDA RESUMEN						
Estadísticas de regresión						
R múltiple	0,9733511429576					
R^2	0,94741244749687					
Error estándar	89,3263755534534					
R^2 ajustado	0,94083900343398					
Observaciones	10					
Análisis de varianza						
	df	SS	MS	F	Significancia de F	
Regresión	1	1150020,3890439	1150020,38904387	144,127254819932	2,136662951644E−06	
Residual	8	63833,610956133	7979,2013695166			
Total	9	1213854				
	Coeficientes	Error estándar	Estadísticas-t	Valor-P	0,95	0,95
Interceptar	−727,7736780964	113,82644006512	−6,393713777573	0,00021053881005	−990,25791958216	−465,2894366107
Columna 1	426,33612821141	35,512322780532	12,0053011132554	2,136662951644E−06	344,444565028797	508,227691394022

a) Si se propone un modelo lineal se obtiene el ajuste: $\hat{y} = -727.774 + 426.336x$.

b) Para el caso del modelo polinomial de orden 2 de la se obtiene l función: $\hat{y} = -704.040 + 409.995x + 2.629x^2$.

c) Para el caso cuadrático se considera $r^2 \approx 0.95$. Para el caso lineal se toma $r \approx 0,97$.

d) Para el modelo lineal simple se tiene:

IC del 95% para α (intersección con el eje y): (-990.258, -465.289) IC del 95% para β (pendiente de la recta): (344.444, 508.228) Para el modelo que contiene el término cuadrático de la variable explicativa: IC del 95% para α (intersección con el eje y): (-2031.383, 623.303) IC del 95% para β_1 (coeficiente del término lineal): (-486.697, 1306.687) IC del 95% para β_2 (coeficiente del término cuadrático): (-140.888, 146.146)

e) El coeficiente correspondiente al término cuadrático no resulta significativamente distinto de 0 (valor p = 0.967). Además se tiene que para ambos modelos $r^2 = 0.947$, lo que indica que la incorporación del término cuadrático no aporta ninguna información para explicar la variabilidad de y si ya incorporamos x al modelo. A su vez para el caso lineal corresponde evaluar el valor de $r = 0,973$ por lo que el ajuste lineal es levemente mejor que el cuadrático.

f) Interpretaciones para conversar en clase.

Media	0,00
Error estándar	26,63
Mediana	−26,31
Moda	#N/D
Desviación estándar	84,22
Varianza de la muestra	7092,62
Curtosis	1,37
Desviación	1,09
Rango	297,90
Mínimo	−120,45
Máximo	177,45
Suma	0,00
Cuenta	10,00