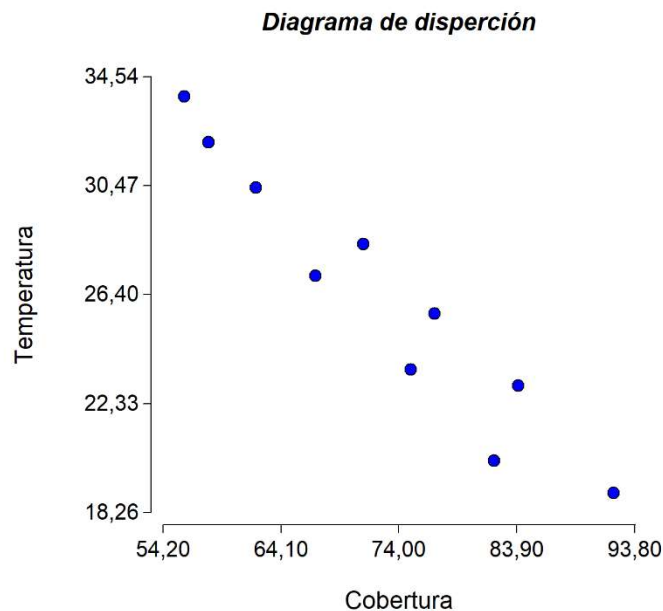


RESOLUCION TP N° 8:

Análisis de Regresión y Correlación

Ejercicio 1: ENSAYO EN VERANO

- a) Variable independiente: % de cobertura arbórea. Es controlada por el investigador y es de carácter fija.
Variable dependiente: Temperatura del suelo. Es la variable aleatoria que depende de la variable independiente y cuyo comportamiento se quiere modelar.
- b) **Diagrama de dispersión:** puede observarse en el gráfico que la nube de puntos podría ajustarse a un modelo lineal con pendiente negativa.



- c) **Estimación de los parámetros de modelo:** aplicando el ajuste por mínimos cuadrados se obtienen las estimaciones de los parámetros del modelo. En este caso el modelo elegido fue la línea recta, que tiene 2 parámetros β_0 y β_1 ; ordenada al origen y pendiente respectivamente.

Análisis de regresión lineal

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Temperatura	10	0,93	0,92	2,92	39,25	40,16

d)

Coefficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI (95%)	LS (95%)	T	p-valor
const	55,52	2,95	48,72	62,33	18,82	<0,0001
Cobertura	-0,40	0,04	-0,50	-0,31	-10,01	<0,0001

c)

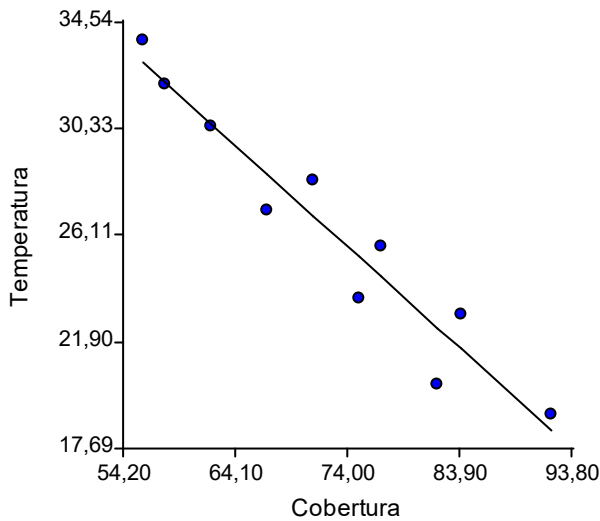
Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	204,09	1	204,09	100,28	<0,0001
Cobertura	204,09	1	204,09	100,28	<0,0001
Error	16,28	8	2,04		
Total	220,38	9			

e)

d)

Diagrama de dispersión



Explicitar el modelo:

Temperatura = 55,52 °c – 0,40 °c/% * % Cobertura

d) Coeficiente de determinación R^2 :

$$R^2 = \frac{SC_{exp}}{SC_{total}}$$

$$R^2 = \frac{204,09}{220,38} = 0,926$$

Este valor nos estaría indicando que el 92,6% de la variabilidad de la temperatura del suelo se puede explicar en este modelo por el % de cobertura arbórea.

- e) **Pruebas de Hipótesis sobre los parámetros:** Debemos realizar las pruebas t para cada uno de los estimadores de los parámetros para conocer si las estimaciones son significativas o no.

Prueba para la Ordenada al origen:

- Planteo de hipótesis:
 $H_0: \beta_0=0$
 $H_1: \beta_0 \neq 0$

$$t_a = \frac{a}{s_a} = \frac{55,52}{2,95} = 18,82$$

Observando su p-valor menor a 0,0001 o comparando contra el t de tabla o teórico (gl=8 y $\alpha=0,05$) de $\pm 2,3$; podemos rechazar la H_0 y concluir que la ordenada al origen no vale 0.

Prueba para la Pendiente:

- Planteo de hipótesis:
 $H_0: \beta_1=0$
 $H_1: \beta_1 \neq 0$

$$t_b = \frac{b}{s_b} = \frac{-0,40}{0,04} = -10,01$$

Observando su p-valor menor a 0,0001 o comparando contra el t de tabla o teórico (gl=8 y $\alpha=0,05$) de $\pm 2,3$; podemos rechazar la H_0 y concluir que la pendiente no vale 0.

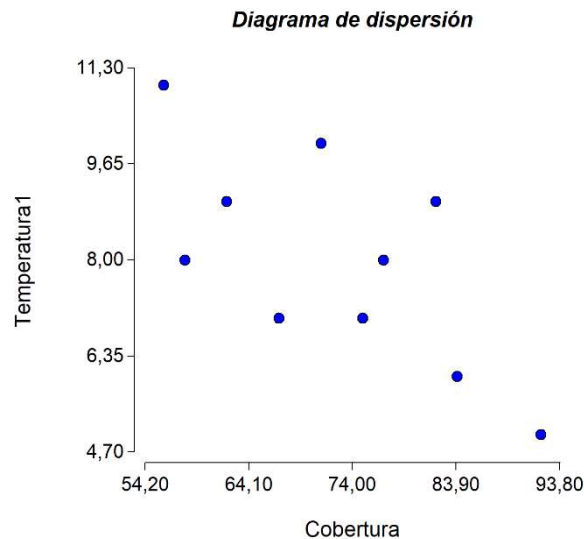
- f) Si necesitamos mantener el cultivo por debajo de los 30° podemos utilizar el modelo que acabamos de estimar para determinar el % de cobertura necesario para mantener esa T° .

$$\begin{aligned} \text{Temperatura} &= 55,52 \text{ }^\circ\text{C} - 0,40 \text{ }^\circ\text{C}/\% * \% \text{ Cobertura} \\ 30^\circ\text{C} &= 55,52^\circ\text{C} - 0,40 \text{ }^\circ\text{C}/\% * \% \text{ Cobertura} \\ \% \text{ Cobertura} &= (30^\circ\text{C} - 55,52^\circ\text{C}) / -0,40 \text{ }^\circ\text{C}/\% \\ \% \text{ Cobertura} &= 63,8 \% \end{aligned}$$

Con un % de cobertura del 63,8 o mayor podemos mantener la temperatura del suelo por debajo de los 30°

Ejercicio 2: ENSAYO EN INVIERNO

- a) Variable independiente: % de cobertura arbórea. Es controlada por el investigador y es de carácter fija.
Variable dependiente: Temperatura del suelo. Es la variable aleatoria que depende de la variable independiente y cuyo comportamiento se quiere modelar.
- b) **Diagrama de dispersión:** puede observarse en el gráfico que la nube de puntos podría ajustarse a un modelo lineal con pendiente negativa.



- c) **Estimación de los parámetros de modelo:** aplicando el ajuste por mínimos cuadrados se obtienen las estimaciones de los parámetros del modelo. En este caso el modelo elegido fue la línea recta, que tiene 2 parámetros β_0 y β_1 ; ordenada al origen y pendiente respectivamente.

Análisis de regresión lineal

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Temperatura	10	0,46	0,40	3,19	39,15	40,05

d)

Coefficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI (95%)	LS (95%)	T	p-valor
const	15,62	2,94	8,85	22,39	5,32	0,0007
Cobertura	-0,11	0,04	-0,20	-0,01	-2,63	0,0303

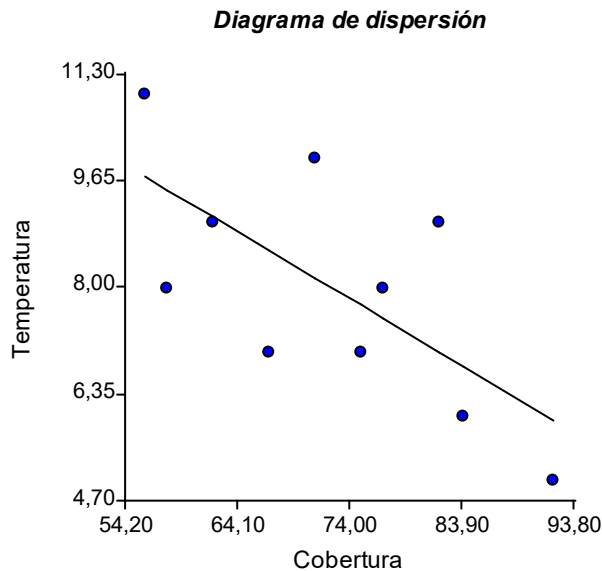
c)

e)

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	13,89	1	13,89	6,90	0,0303
Cobertura	13,89	1	13,89	6,90	0,0303
Error	16,11	8	2,01		
Total	30,00	9			

d)



Explicitar el modelo:

Temperatura = 15,62 °c – 0,11 °c/% * % Cobertura

d) Coeficiente de determinación R^2 :

$$R^2 = \frac{SC_{exp}}{SC_{total}}$$

$$R^2 = \frac{13,89}{30} = 0,463$$

Este valor nos estaría indicando que el 46,3% de la variabilidad de la temperatura del suelo se puede explicar en este modelo por el % de cobertura arbórea.

e) **Pruebas de Hipótesis sobre los parámetros:** Debemos realizar las pruebas t para cada uno de los estimadores de los parámetros para conocer si las estimaciones son significativas o no.

Prueba para la Ordenada al origen:

- Planteo de hipótesis:
H₀: β₀=0
H₁: β₀≠0

$$t_a = \frac{a}{s_a} = \frac{15,62}{2,94} = 5,32$$

Observando su p-valor menor a 0,0007 o comparando contra el t de tabla o teórico (gl=8 y α=0,05) de ±2,3; podemos rechazar la H₀ y concluir que la ordenada al origen no vale 0.

Prueba para la Pendiente:

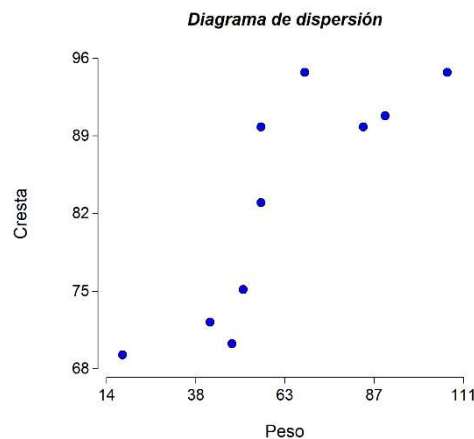
- Planteo de hipótesis:
H₀: β₁=0
H₁: β₁≠0

$$t_b = \frac{b}{s_b} = \frac{-0,11}{0,04} = -2,63$$

Observando su p-valor menor a 0,0303 o comparando contra el t de tabla o teórico (gl=8 y α=0,05) de ±2,3; podemos rechazar la H₀ y concluir que la pendiente no vale 0.

Comparación con Ejercicio 1: al tratarse de una especie caducifolia y comparar la T° del suelo en función del % de cobertura, es lógico encontrar diferencias entre invierno y verano. El modelo de verano es mucho más significativo y con mayor variabilidad explicada (R²) que el modelo de invierno.

Ejercicio 3: Diagrama de dispersión entre el peso en gr y la longitud en cm de la cresta de 10 pollitos blancos de la raza Leghorn.



Covarianza:

$$\text{Covarianza}(xy) = \sum_{i=1}^n X \cdot Y - \frac{\sum_{i=1}^n X \cdot \sum_{i=1}^n Y}{n}$$

	X	Y	X.Y
	83	56	4648
	72	42	3024
	69	18	1242
	90	84	7560
	90	56	5040
	95	107	10165
	91	90	8190
	95	68	6460
	75	51	3825
	70	48	3360
Suma	830	620	53514

$$\text{Covarianza}(xy) = \sum_{i=1}^n X \cdot Y - \frac{\sum_{i=1}^n X \cdot \sum_{i=1}^n Y}{n} = 53514 - \frac{830 \cdot 620}{10} = 2054$$

Coefficiente de correlación (r):

$$r = \frac{\sum_{i=1}^n X \cdot Y - \frac{\sum_{i=1}^n X \cdot \sum_{i=1}^n Y}{n}}{\sqrt{\left[\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right] \cdot \left[\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right]}}$$

	X	Y	X.Y	X²	Y²
	83	56	4648	6889	3136
	72	42	3024	5184	1764
	69	18	1242	4761	324
	90	84	7560	8100	7056
	90	56	5040	8100	3136
	95	107	10165	9025	11449
	91	90	8190	8281	8100
	95	68	6460	9025	4624
	75	51	3825	5625	2601
	70	48	3360	4900	2304
Suma	830	620	53514	69890	44494

$$r = \frac{53514 - \frac{830 \cdot 620}{10}}{\sqrt{\left[69890 - \frac{(830)^2}{10}\right] \cdot \left[44494 - \frac{(620)^2}{10}\right]}} = \frac{2054}{2460,49} = 0,83$$

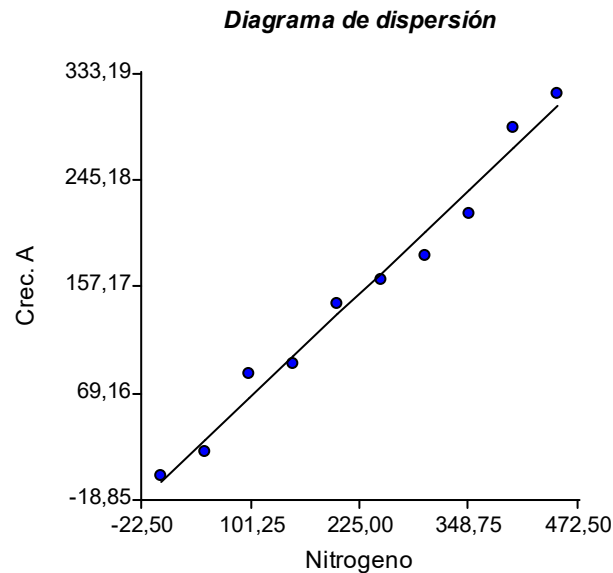
Coefficiente de correlación (r) :

	Peso	Cresta
Peso	1,00	2,7E-03
Cresta	0,83	1,00

En este caso el coeficiente de correlación (r) = 0,83 es un valor alto (cercano a +1) y significativamente distinto de 0 (p-valor = $2,7e^{-03}$); por lo que podemos concluir que las variables están correlacionadas positivamente.

Ejercicio 4:

Especie A: Crecimiento = -3,37 mg/día + 0,69 mg/día / kg/ha * kg/ha Nitrógeno



Análisis de regresión lineal

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Crec. A	10	0,98	0,98	357,88	86,53	87,43

Coefficientes de regresión y estadísticos asociados

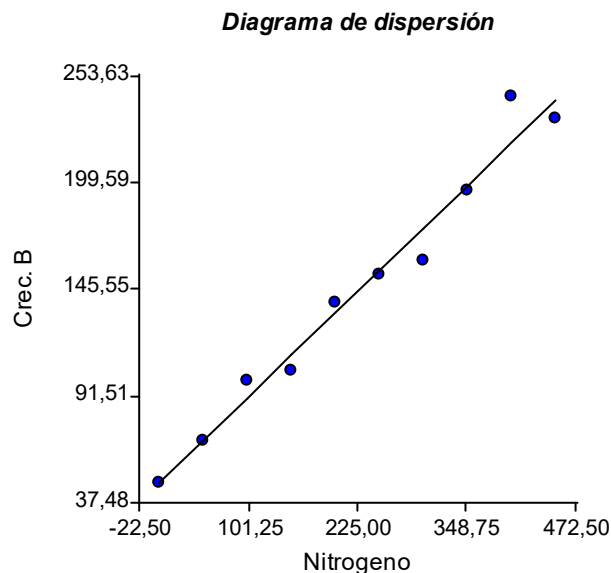
Coef	Est.	E.E.	LI (95%)	LS (95%)	T	p-valor
const	-3,37	8,91	-23,92	17,19	-0,38	0,7156
Nitrogeno	0,69	0,03	0,61	0,77	20,60	<0,0001

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	97634,08	1	97634,08	424,55	<0,0001
Nitrogeno	97634,08	1	97634,08	424,55	<0,0001
Error	1839,78	8	229,97		
Total	99473,86	9			

Analizando a la Especie A vemos que su ordenada al origen no es significativamente diferente de 0. Entonces podemos pensar que dicha especie sin la disponibilidad de Nitrógeno no crecerá y ante la disponibilidad de Nitrógeno lo hará a una tasa de 0,69 mg/día (ya que la pendiente si es significativa). El Nitrógeno explica el 98 % del crecimiento en esta especie.

Especie B: Crecimiento = 47,41 mg/día + 0,43 mg/día / kg/ha * kg/ha Nitrógeno



Análisis de regresión lineal

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
Crec. B	10	0,97	0,97	227,57	81,54	82,45

Coefficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI (95%)	LS (95%)	T	p-valor
const	47,41	6,95	31,39	63,43	6,82	0,0001
Nitrogeno	0,43	0,03	0,37	0,49	16,56	<0,0001

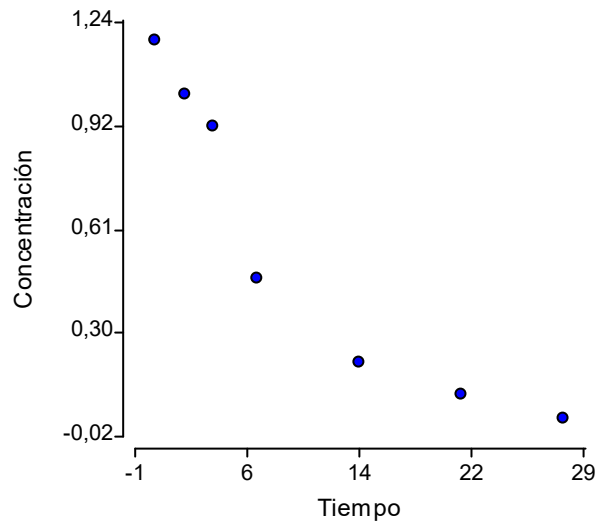
Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	38327,22	1	38327,22	274,30	<0,0001
Nitrogeno	38327,22	1	38327,22	274,30	<0,0001
Error	1117,82	8	139,73		
Total	39445,04	9			

Analizando a la Especie B vemos que su ordenada al origen si es significativamente diferente de 0. Entonces podemos pensar que dicha especie ante la ausencia de Nitrógeno crecerá 47,41 mg/día y ante la disponibilidad de Nitrógeno lo hará a una tasa de 0,43 mg/día (ya que la pendiente también es significativa). El Nitrógeno explica el 97 % del crecimiento en esta especie.

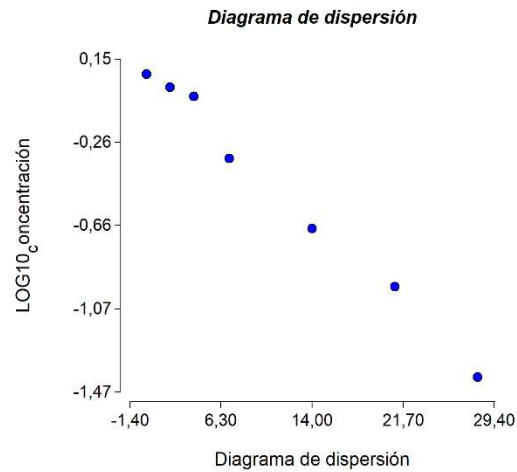
Ejercicio 5: Concentración de pesticida en función del tiempo.

Diagrama de dispersión



Al observar el gráfico de dispersión vemos que la nube de puntos sigue una forma curva por lo que descartamos un ajuste lineal directamente. Vamos a intentar linealizar la curva aplicando una transformación logarítmica a las variables. Al aplicarle logaritmo a la variable Y lo que se intenta es linealizar un modelo exponencial.

$$\text{Concentración} = a * e^{b * \text{Tiempo}}$$



$$\text{LOG Concentración} = a' + b * \text{Tiempo}$$

Análisis de regresión lineal

Variable	N	R ²	R ² Aj	ECMP	AIC	BIC
LN Concentración	7	0,99	0,99	0,03	-5,23	-5,39

Coefficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI (95%)	LS (95%)	T	p-valor
const	0,23	0,07	0,04	0,41	3,11	0,0264
Tiempo	-0,12	5,0E-03	-0,13	-0,11	-24,43	<0,0001

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo	9,84	1	9,84	596,70	<0,0001
Tiempo	9,84	1	9,84	596,70	<0,0001
Error	0,08	5	0,02		
Total	9,92	6			

$$\text{LOG Concentración} = 0,23 - 0,12 * \text{Tiempo}$$

c) Estimar

A- T₂ si m=2 entonces Conc= C₀ / m → Conc=1,18/2 = 0,59

Log(0,59)= -0,229

$$\text{Tiempo}_2 = (-0,229 - 0,23) / -0,12$$

$$\text{Tiempo}_2 = 3,825 \text{ días}$$

B- T_{10} si $m=10$ entonces $\text{Conc} = C_0 / m$

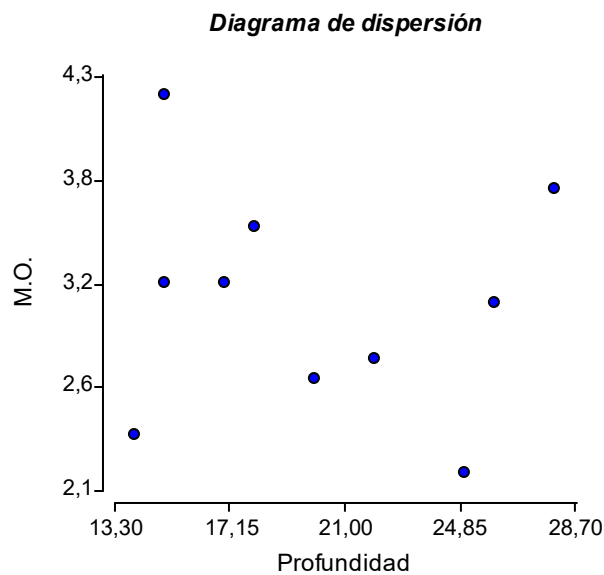
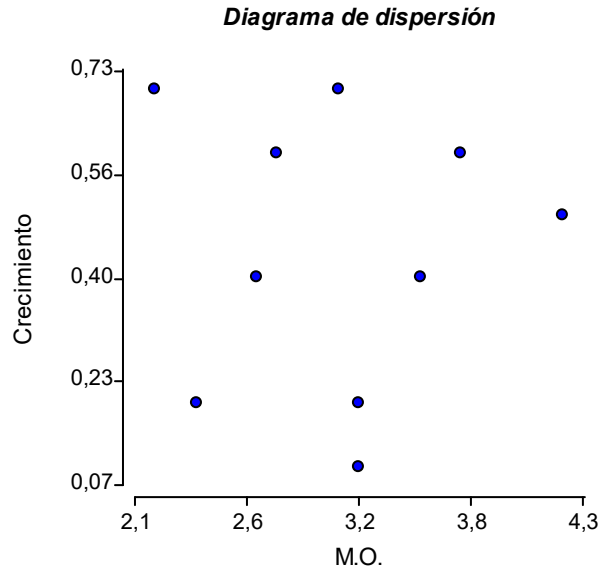
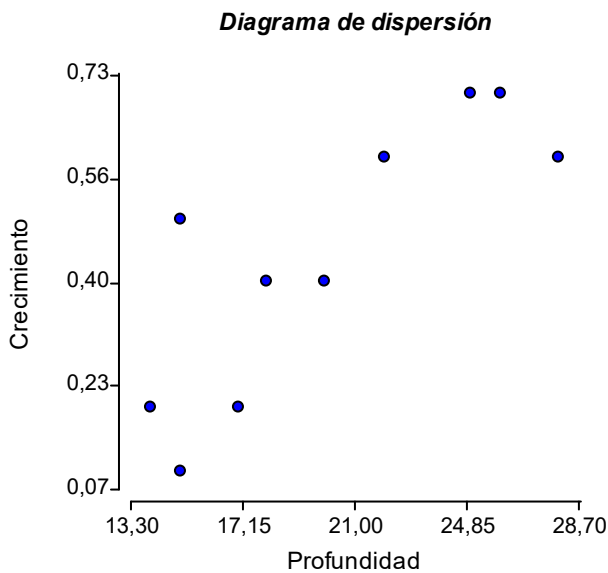
$$\text{Conc} = 1,18 / 10 = 0,118$$

$$\text{Log}(0,118) = -0,928$$

$$\text{Tiempo}_{10} = (-0,928 - 0,23) / -0,12$$

$$\text{Tiempo}_{10} = 9,65 \text{ días}$$

Ejercicio 6: Crecimiento de una especie arbórea relacionada a las características del suelo (profundidad y materia orgánica)



- a) Observando los diagramas de dispersión para cada par de variables podemos apreciar que podría existir una relación lineal entre Crecimiento y Profundidad mientras que no se observa relación aparente entre Crecimiento y Materia Orgánica ni entre Profundidad y Materia Orgánica.
- b) Matriz de correlaciones: debajo de la diagonal se encuentran los valores de los coeficientes de correlación y sobre ella se encuentran los p-valores de las pruebas de hipótesis para cada coeficiente.

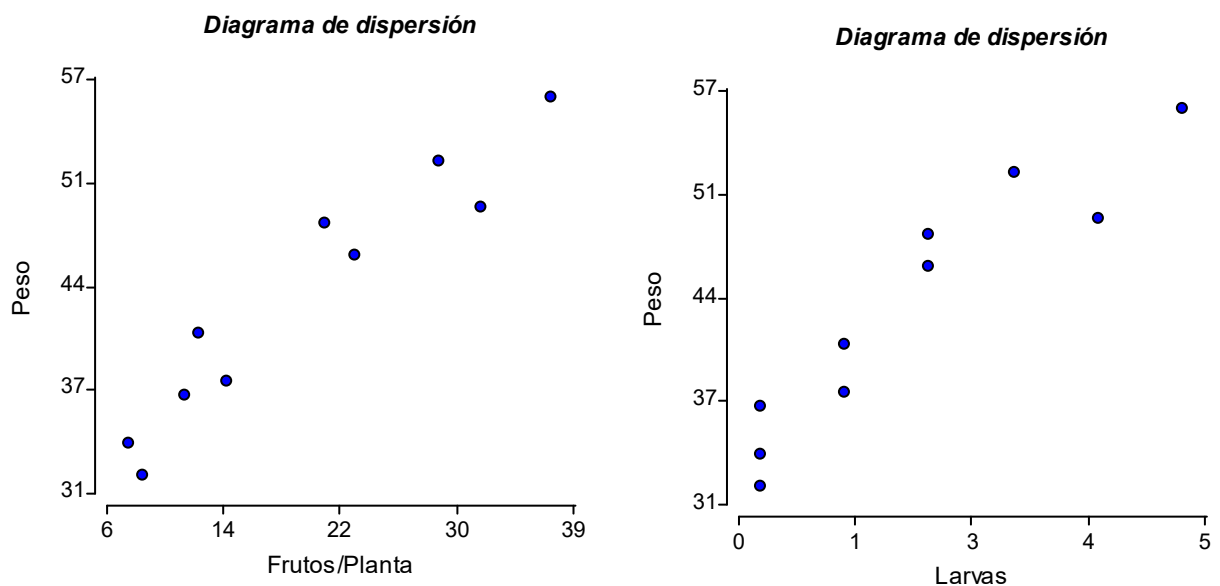
Coeficientes de correlación

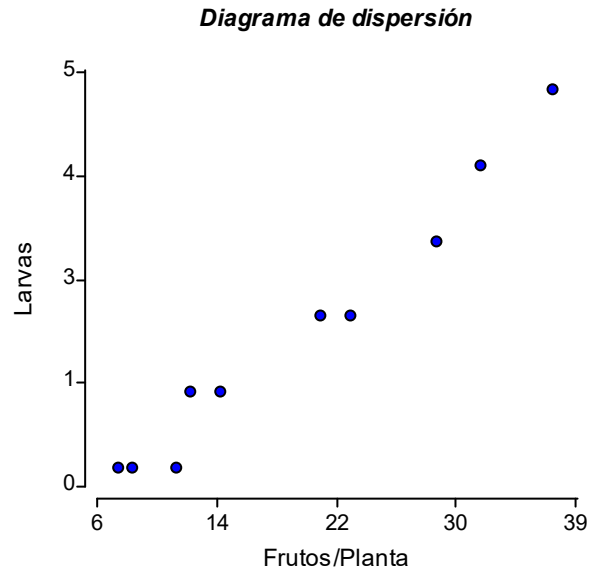
Correlacion de Pearson: coeficientes\probabilidades

	Crecimiento	Profundidad	M.O.
Crecimiento	1,00	3,4E-03	0,98
Profundidad	0,82	1,00	0,75
M.O.	-0,01	-0,12	1,00

- c) , d) y e) La única relación significativa es la del Crecimiento con la Profundidad ($r=+0,82$); el resto de los coeficientes no son significativamente diferentes de 0.
La relación del crecimiento de esta especie con la profundidad es positiva y alta, por lo que a mayor Profundidad mayor será el Crecimiento. La Materia orgánica no tendría relación ni influencia en el Crecimiento.

Ejercicio 7: Relación entre el tamaño y número de frutos con el número de insectos plaga.





a) Observando los diagramas de dispersión para cada par de variables podemos apreciar que podría existir una relación lineal entre todas ellas.

b) Matriz de correlaciones:

Coeficientes de correlación

Correlación de Pearson: coeficientes\probabilidades

	Larvas	Peso	Frutos/Planta
Larvas	1,00	5,2E-05	2,2E-07
Peso	0,94	1,00	1,3E-05
Frutos/Planta	0,98	0,96	1,00

c) El Peso del fruto y número de larvas están relacionados positivamente ($r=+0,94$); a mayor peso del fruto mayor el número de larvas que se espera encontrar.

El Número de frutos por planta y el número de larvas están relacionados positivamente ($r=+0,98$); a mayor cantidad de frutos por planta mayor el número de larvas que se espera encontrar.

d) El Peso del fruto y el Número de frutos por planta también se relacionan positivamente ($r=+0,96$) por lo que cuanto más frutos tiene una planta más pesados serán.

e) Todos los coeficientes resultaron ser altamente significativos; por lo que se entiende que todas las variables están correlacionadas entre sí.