Análisis de Regresión Lineal

Erika Kania Kuhl
Ing. Agrónomo Dr.

Modelo de regresión lineal

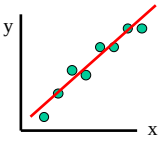
"Técnica estadística cuyo fundamento es la construcción de un "*modelo*" que permite la estimación de la media de una variable llamada *dependiente* para valores determinados de una variable o varias variables *independientes* o *regresoras*"

Variable dependiente "Y" variables independientes "X"

Modelo de regresión lineal simple: Una variable dependiente y una variable independiente

Modelo de regresión lineal

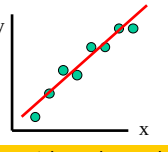
El análisis de regresión sirve para **predecir** una variable dependiente en función de una (RLS) o varias (RLM) variables independientes



Modelo de regresión lineal

El análisis de regresión sirve para **predecir** una variable dependiente en función de una (RLS) o varias (RLM) variables independientes

Y = Variable dependiente
Predicha
Explicada
Respuesta



x = Variable independiente
Predictora
Explicativa
Regresora
Causal

Modelo de regresión lineal

El análisis de regresión sirve para **predecir** una variable dependiente en función de una (RLS) o varias (RLM) variable independiente

Predecir =
- Relación.....
- Función.....
- Dependencia.....
- Efecto causa / respuesta
- Estimación

Modelo de regresión lineal

El análisis de regresión sirve para **predecir** una variable dependiente en función de una (RLS) o varias (RLM) variable independiente

Predecir =
- Relación.....
- Función.....
- Dependencia.....
- Efecto causa / respuesta
- Estimación

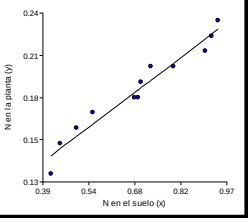
Asociación….

- Análisis de Correlación lineal
- Asociación entre variables
- Coeficiente de correlación lineal de Pearson (r)

Otro capítulo : Correlación lineal

Modelo de regresión lineal

1. **Identificar** a la variable dependiente y a la (s) variable (s) independientes
2. **Graficar** la relación entre las variables
3. **Identificar el modelo** que relaciona las variables
4. **Estimar los parámetros** del modelo
5. **Probar hipótesis** sobre los parámetros del modelo
6. **Predecir** el nivel medio de la variable respuesta para los valores determinados de la (s) variables independientes



Ejemplos de relaciones funcionales de interés Agronómico:

•Relación entre el rendimiento de un cultivo y la densidad de siembra

•Relación entre la cantidad de suplemento dado y el aumento de peso que éste produce en un lote de animales

•Relación entre las dosis de un insecticida y la mortalidad de los insectos tratados

• Relación entre ASTT versus producción por planta

Identificar la variable dependiente y la variable independiente

Modelo de regresión lineal simple

Modelo teórico:

$$y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

Y ij= observación de la **variable dependiente** bajo el i-ésimo nivel de X, i = 1,...,K en la j-ésima unidad experimental, j = 1,...,m
X i= i-ésimo valor de la **variable independiente**, i = 1,...,K
α = parámetro que representa la ordenada al origen de la recta (indica valor esperado de Y cuando X=0)
β = parámetro que representa la pendiente de la recta (tasa de cambio en Y frente al cambio unitario en X).
εij = variación aleatoria (o no explicada por el modelo) asociada a la j-ésima observación de Y bajo el nivel Xi.

Los εij se suponen normales e independientemente distribuidos con esperanza 0 y varianza constante σ2 para todo X en un intervalo donde el modelo se supone verdadero. Esto es εij ~ N I D (0, σ2)

Modelo de regresión lineal simple. Modelo teórico

$$y_{ij} = \beta_0 + \beta_1 X_1 + \varepsilon_{ij}$$

Modelo de regresión lineal múltiple. Modelo teórico

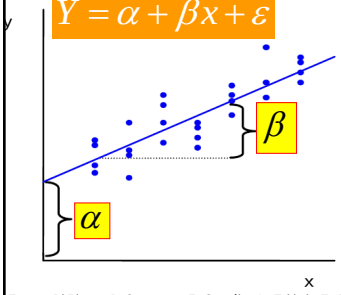
$$y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \ldots + \beta_{p-1} X_{p-1} + \varepsilon_{ij}$$

Representación de los datos

- La manera de mostrar gráficamente los datos observados en un gráfico es a través de un diagrama de dispersión
- Y, la respuesta se marca en el eje vertical
- X, la variable explicativa, en el eje horizontal
- Cada observación es un punto del gráfico

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

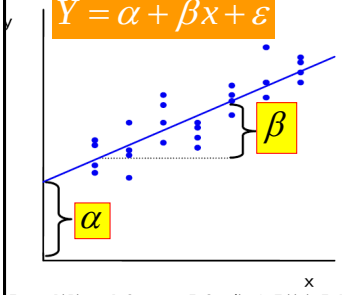
$$Y = \alpha + \beta x + \varepsilon$$



Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

$$Y = \alpha + \beta x + \varepsilon$$

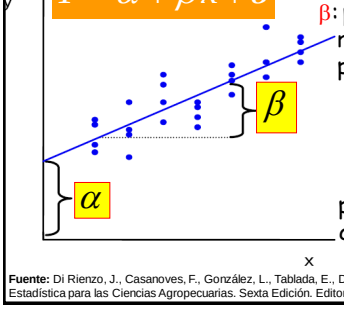


α : parámetro que representa la ordenada al origen de la recta (indica el valor esperado de Y cuando x = 0)

Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

$$Y = \alpha + \beta x + \varepsilon$$



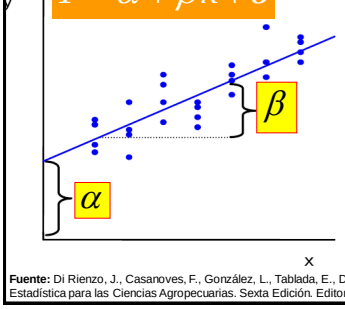
β: parámetro que representa la pendiente de la recta

(Su magnitud sirve para predecir en cuanto aumentará en promedio "y" cada vez que "x" se incremente en una unidad)

Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

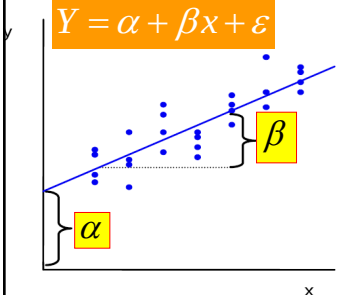
$$Y = \alpha + \beta x + \varepsilon$$



Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

$$Y = \alpha + \beta x + \varepsilon$$



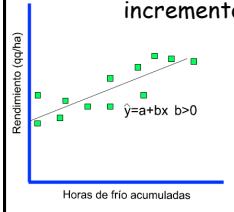
Luego, la pendiente y la ordenada al origen determinan la posición de la recta.

Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

$\beta > 0$

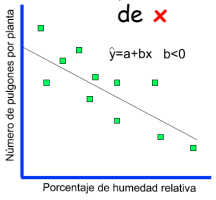
Significa que hay un crecimiento de β unidades en y (en promedio) por cada incremento de una unidad en x



Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

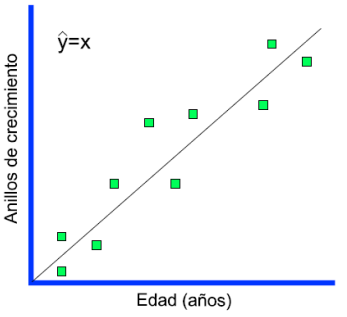
¿Cómo se interpretan los parámetros del modelo de regresión lineal simple ?

$\beta < 0$ Significa que **y** disminuirá β unidades (en promedio) con cada incremento unitario de **x**



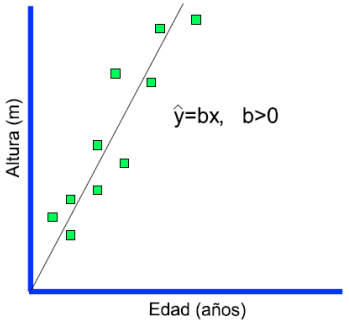
Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Gráficos de dispersión



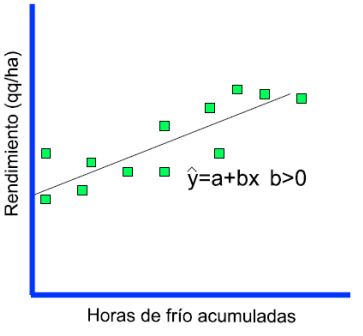
Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Gráficos de dispersión



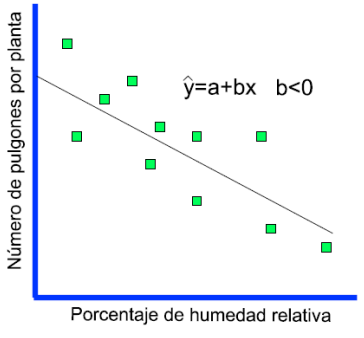
Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Gráficos de dispersión



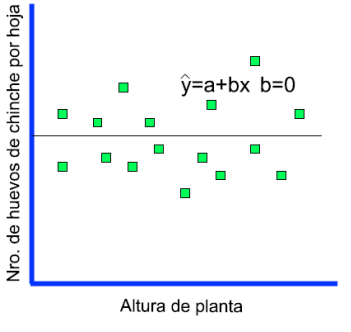
Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Gráficos de dispersión



Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Gráficos de dispersión



Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Archivo Reg Trigo

En un ensayo sobre trigo se desea cuantificar la relación que hay entre la disponibilidad de nitrógeno en el suelo y la cantidad de nitrógeno en la planta

Se obtuvieron datos para 12 parcelas, en las que se registró el contenido de nitrógeno en el suelo (x) y los valores promedios de nitrógeno por planta (y).

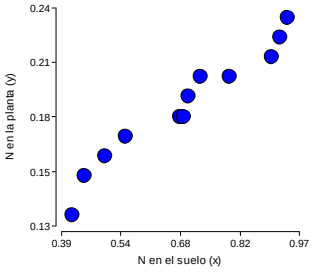
Ejemplo

X: Nitrógeno en el suelo (ppm)	Y: Nitrógeno en la planta (ppm)
0,42	0,13
0,45	0,15
0,50	0,16
0,55	0,17
0,68	0,18
0,69	0,18
0,70	0,19
0,73	0,20
0,80	0,20
0,90	0,21
0,92	0,22
0,94	0,23

Cada fila representa los valores observados sobre la unidad experimental, conformada por una parcela de 50 cm. x 50 cm., en la que se midió el nitrógeno en el suelo y por planta calculado como promedio sobre todas las plantas de la parcela

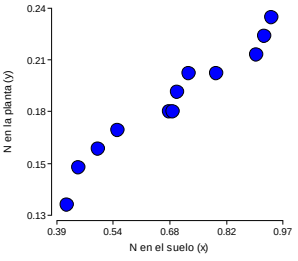
Ejemplo

El diagrama de dispersión para los datos se presenta en la siguiente figura:



Ejemplo

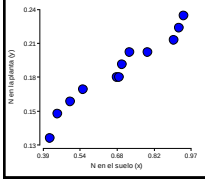
¿Que se desprende de la observación del gráfico?



Ejemplo

¿Que se desprende de la observación del gráfico?

El diagrama indica que hay una relación positiva entre la cantidad de nitrógeno en la planta y la cantidad de nitrógeno disponible en el suelo.

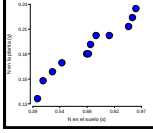


En este ejemplo se puede postular una relación lineal.

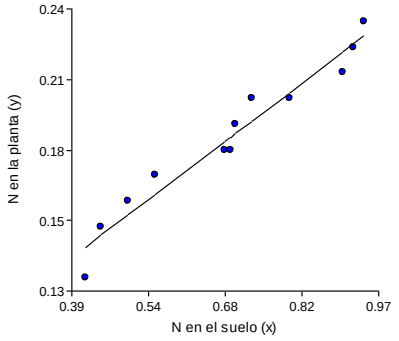
Ejemplo

Si se usa la regresión para hacer la predicción de "y" a partir de "x", el objetivo será trazar la línea que mejor se ajuste a los puntos.

Esa recta hace una predicción de qué valores irá tomando "y" (N en la planta) en función de "x" (N en el suelo).



Ejemplo



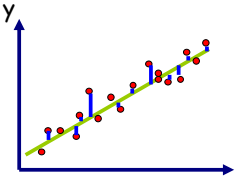
Error

Error

Se aprecia que la recta resume relativamente bien los puntos, pero casi ninguno de los puntos está exactamente sobre ella.

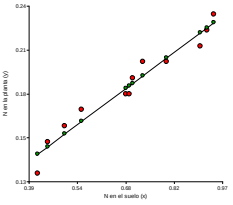
La distancia entre cada punto observado y la recta de regresión es el Residuo (Error, Residual) para cada punto.

Esta distancia expresa el ruido o error que existe en el modelo.



Estimación de la recta de regresión

Si observamos el gráfico de los datos, resulta obvio que no se puede construir ninguna recta que pase a través de todos los puntos.



No importa que recta construyamos, diversos puntos se desviarán de dicha recta.

Estimación de la recta de regresión

Hasta ahora hemos hablado de predicciones a partir de una ecuación de regresión, pero….

¿Cuál es el modo de saber cuáles son los coeficientes que definen la recta que mejor se ajusta a la nube de puntos?

Esto supone saber cuánto valen la ordenada en el origen "a" y la pendiente de la recta "b"

Estimación de la recta de regresión

La ecuación de la recta de regresión es:

$$y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

Estimación de la recta de regresión

La ecuación de la recta de regresión es:

$$y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

Los parámetros teóricos (α y β) hay que estimarlos

Se tiene así la recta ajustada

α

$\hat{y} = a + b\,x$

$\hat{y}$ = y estimada

Estimación de la recta de regresión

$$y_{ij} = \alpha + \beta X_i + \varepsilon_{ij}$$

$\hat{y} = a + b\,x$

α y β: letras griegas que hay que estimar (no los conocemos):

a y **b**: valores conocidos (estimaciones de α y β)

x e **y**: son los valores experimentales

Estimación de la recta de regresión

$\hat{y} = a + b\,x$

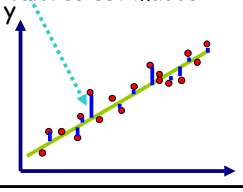
Para estimar o calcular "a" y "b" se usa el
Método de los mínimos cuadrados
(MMC)

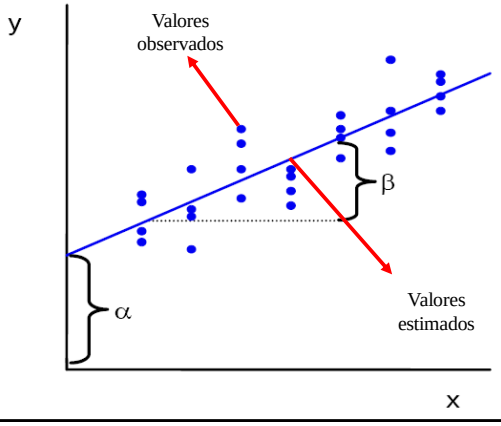
Ajuste de una recta por el Método de los Mínimos Cuadrados (M.M.C)

Estimación de la recta de regresión

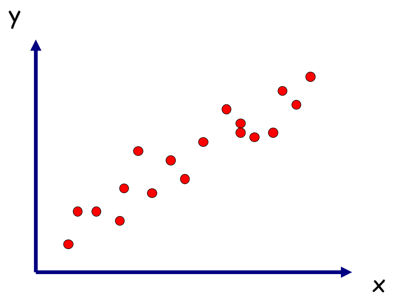
La idea central del Método de los Mínimos Cuadrados (M.M.C.), es estimar los parámetros de tal forma que los RESIDUOS sean lo más pequeño posible.

Es decir, minimizar las distancias entre los valores observados (Yi), y los valores estimados de la recta (y estim).

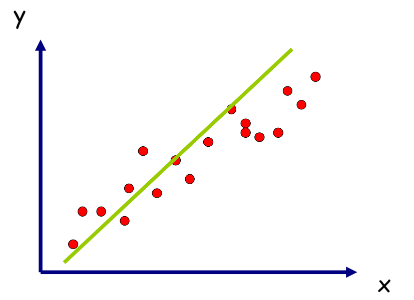




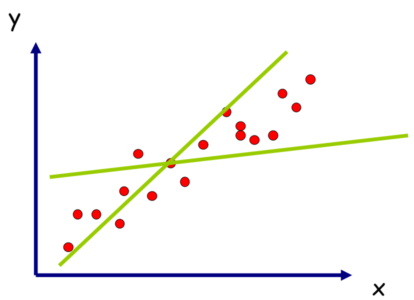
Estimación de la recta de regresión



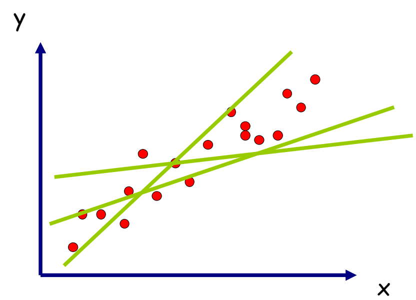
Estimación de la recta de regresión



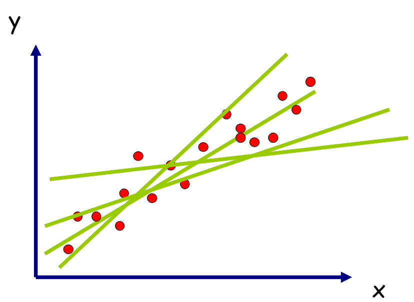
Estimación de la recta de regresión



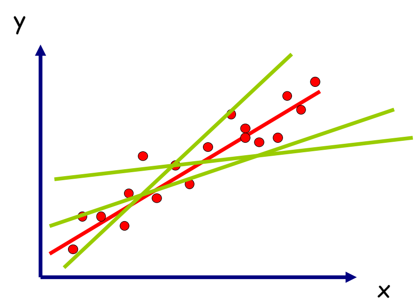
Estimación de la recta de regresión



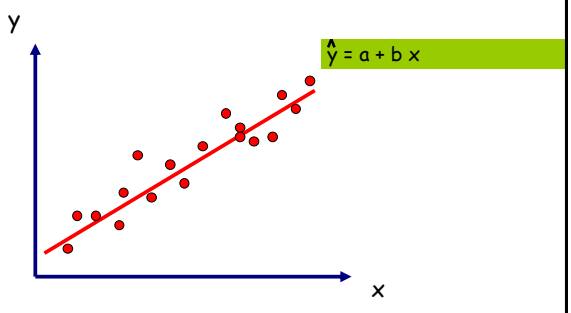
Estimación de la recta de regresión



Estimación de la recta de regresión

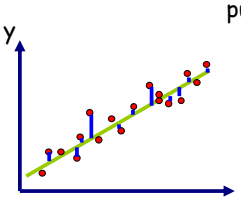


Estimación de la recta de regresión



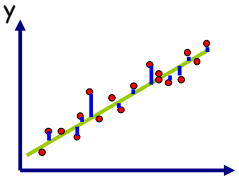
Estimación de la recta de regresión

El **Métodos de mínimos cuadrados** define la recta de "mejor ajuste" (entre todas las posibles) como aquella que minimiza la desviación del "total" de observaciones con respecto a la línea recta (la mas cercana posible a todos los puntos)



Estimación de la recta de regresión

Utilizando esta medida como el criterio para la exactitud de ajuste, trataremos de encontrar la línea recta que hará la suma de cuadrados de las desviaciones tan pequeña como sea posible.



Estimación de la recta de regresión

Por un proceso de minimización de la suma de cuadrados del error, se obtienen los estimadores "a" y "b" de α y β respectivamente, llamados estimadores mínimos cuadráticos, los que se calculan mediante las siguientes formulas:

Estimación de la recta de regresión

$a = \bar{y} - b\,\bar{x}$ (α estimado)

$$b = \frac{\Sigma xy - \Sigma x \Sigma y / n}{\Sigma x^2 - (\Sigma x)^2 / n}$$ (β estimado)

n = número de pares de datos

Se tiene así la recta ajustada:

$\hat{y} = a + b\,x$

Resumiendo

POBLACION	MUESTRA

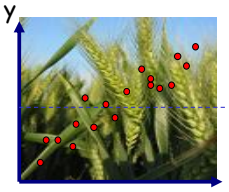
a, b α, β, parámetros, estimadores, letras griegas, letras latina

Resumiendo

POBLACION	MUESTRA
Parámetros	Estimadores
Letras griegas	Letras latinas
α	a
β	b

Parámetros: valor supuesto de una población
Estimadores: valor numérico calculado sobre una muestra

Ejemplo



X: Nitrógeno en el suelo (ppm)	Y: Nitrógeno en la planta (ppm)
0,42	0,13
0,45	0,15
0,50	0,16
0,55	0,17
0,68	0,18
0,69	0,18
0,70	0,19
0,73	0,20
0,80	0,20
0,90	0,21
0,92	0,22
0,94	0,23

Ejemplo: recta estimada

$\hat{y} = a + b\,x$

$\hat{y} = 0.076 + 0.159\,x$

N planta estim. = 0.076 + 0.159 N suelo

a= 0.076 (0.0756284)

b= 0.159 (0.1585094)

Ejemplo: recta estimada

$\hat{y} = a + b\,x$

$\hat{y} = 0.076 + 0.159\,x$

N planta estim. = 0.076 + 0.159 N suelo

¿ Por cada aumento en un ppm de N en el suelo (x) el N en la planta (y) aumenta en promedio en _____ ppm?

Ejemplo: recta estimada

y estimada = 0.076 + 0.159 x

Si:

X=	Y =
1	
2	
3	
4	

Ejemplo: recta estimada

y estimada = 0.076 + 0.159 x

Si:

X=	Y =
1	0.235
2	0.394
3	0.553
4	0.712

Ejemplo: recta estimada

y estimada = 0.076 + 0.159 x

Si:

X=	Y =
1	0.235
2	0.394
3	0.553
4	0.712

0.159

Ejemplo: recta estimada

y estimada = 0.076 + 0.159 x

Si:

X=	Y =
1	0.235
2	0.394
3	0.553
4	0.712

0.159

0.159

Ejemplo: recta estimada

y estimada = 0.076 + 0.159 x

Si:

X=	Y =
1	0.235
2	0.394
3	0.553
4	0.712

0.159

0.159

0.159

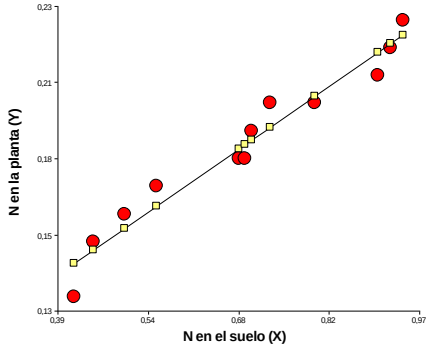
Ejemplo en Infostat

InfoStat/P - Nueva tabla - [Nueva tabla]

Archivo Edición Datos Resultados Estadísticas Gráficos Ventanas Aplicaciones Ayuda [R]

Caso	N en el suelo (X)	N en la planta (Y)	RDUO_N en la planta (Y)	PRED_N en la planta (Y)
1	0,42	0,13	-0,01	0,14
2	0,45	0,15	0,00	0,15
3	0,50	0,16	0,01	0,15
4	0,55	0,17	0,01	0,16
5	0,68	0,18	0,00	0,18
6	0,69	0,18	-0,01	0,19
7	0,70	0,19	0,00	0,19
8	0,73	0,20	0,01	0,19
9	0,80	0,20	0,00	0,20
10	0,90	0,21	-0,01	0,22
11	0,92	0,22	0,00	0,22
12	0,94	0,23	0,01	0,22

Ejemplo en Infostat



Ejemplo en Infostat

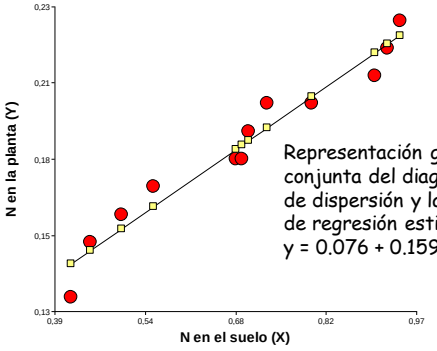
Análisis de regresión lineal

Variable	N	R²	R² Aj	ECMP	AIC	BIC
N en la planta (Y)	12	0,951	0,946	7,4E-05	-81,847	-80,392

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor
const	0,076	0,008	0,058	0,094	9,349	<0,0001
N en el suelo (X)	0,159	0,011	0,133	0,184	13,940	<0,0001

Ejemplo en Infostat



Representación gráfica conjunta del diagrama de dispersión y la recta de regresión estimada
y = 0.076 + 0.159 x

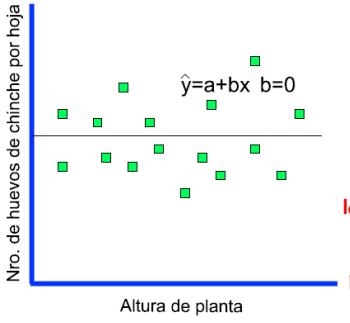
Pruebas de hipótesis para la pendiente

Para probar si la recta de regresión es un modelo adecuado para expresar la relación lineal entre la respuesta (variable dependiente) "Y" y la variable predictora (variable independiente) "X", en el tramo investigado, o lo que es lo mismo, si la recta se ajusta relativamente bien a la nube de puntos (linealmente) debe hacerse mediante una Prueba de Hipótesis.

Pruebas de hipótesis para la pendiente

Ho : β = 0
Si aceptamos Ho, **la conclusión** es:
No existe efecto lineal de "x" sobre "y" o la pendiente es constante
El modelo no es significativo

H_A : β ≠ 0
Si rechazamos Ho, **la conclusión** es:
Existe efecto lineal de "x" sobre "y" o la pendiente no es constante.
El modelo es significativo



Ho : β = 0

(No existe efecto lineal de "x" sobre "y")

No importa cual sea la altura de la planta, el número de huevos de chinche por hoja permanece constante

Fuente: Di Rienzo, J., Casanoves, F., González, L., Tablada, E., Díaz, M., Robledo, C. y Balzarini, M. 2005. Estadística para las Ciencias Agropecuarias. Sexta Edición. Editorial Brujas. 329 p.

Pruebas de hipótesis para la pendiente

Ho : β = 0

H_A : β ≠ 0

Para probar estas hipótesis en el caso de la **regresión lineal simple**, se puede realizar por dos métodos equivalentes:

1. Mediante una **Prueba F** a través del ANDEVA
2. Mediante una prueba **t de Student**

Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP	AIC	BIC
N en la planta (Y)	12	0,951	0,946	7,4E-05	-81,847	-80,392

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	0,076	0,008	0,058	0,094	9,349	<0,0001		
N en el suelo (X)	0,159	0,011	0,133	0,184	13,940	<0,0001	177,744	1,000

Cuadro de Análisis de la Varianza (SC tipo III)

F.V.	SC	gl	CM	F	p-valor
Modelo.	0,009	1	0,009	194,318	<0,0001
N en el suelo (X)	0,009	1	0,009	194,318	<0,0001
Error	4,6E-04	10	4,6E-05		
Total	0,010	11			

Coeficiente de determinación = R^2

•Se interpreta como el porcentaje de la variabilidad total de la variable dependiente ("y") que es explicado por la variable independiente ("x") (por la regresión, por el modelo).

Coeficiente de determinación = R^2

$0 \leq R^2 \leq 1$

$0 \leq R^2 \leq 100$ %

R^2 alto. Indica que hay poca variabilidad de las observaciones con respecto a esa recta ajustada

R^2 bajo. Indica que hay mucha variabilidad de las observaciones con respecto a esa recta ajustada

Coeficiente de determinación = R^2

¿Que significa que R^2 = 95.1 %?

Quiere decir que la variable independiente "x" explica un 95,1 % de la respuesta "y" y por lo tanto queda un 4,9 % de ruido o error.

Es decir, el contenido de N en el suelo explica un 95,1 % de la variación total del contenido de N en la planta (a través de una relación lineal). El 4,9 % restante es ruido.

Coeficiente de determinación R^2

Indica como es la variabilidad de las observaciones con respecto a esa recta ajustada

El R^2 lo único que nos dice es que tan disperso están los datos con respecto a la recta.

R^2 no indica si un modelo es mejor o peor que otro.

R^2 no es un método de bondad de ajuste de modelo.

No se pueden elegir modelos comparando el R^2

Mientras más parámetros agregamos al modelo siempre aumenta el R^2

Interpretación de α

Es la respuesta promedio (valor promedio de Y) en ausencia de estimulo (x = 0), siempre que en el ajuste de los datos de hayan utilizado valores de X cercanos a cero o que este valor sea extrapolable

Interpretación del intercepto

El valor del "intercepto" es un ajuste matemático, pero biológicamente no nos dice nada.

Diámetro bayas = 15,3 + 0,9 ddpf

Interpretación del intercepto

El valor del "intercepto" es un ajuste matemático, pero biológicamente no nos dice nada.

Diámetro bayas = 15,3 + 0,9 ddpf

El valor del intercepto es 15,3, matemáticamente estaría indicando que a los 0 ddpf (días después de plena flor) el diámetro promedio de las bayas de vid es 15,3, lo que sería imposible por que en plena flor no existen bayas cuajadas

Ajuste con fines predictivos

En múltiples ocasiones se desea utilizar el ajuste lineal con fines predictivos, es decir obtener los valores estimados promedio de "y" para un valor de "x". Se pueden dar dos situaciones:

1) Para un valor de "x" que esté dentro del rango utilizado en el ajuste (INTERPOLACIÓN)

En este caso se puede hacer una "predicción" siempre que se pruebe que la recta presenta un buen ajuste

Ajuste con fines predictivos

2) Para un valor de "x" que esté fuera del rango utilizado en el ajuste (EXTRAPOLACIÓN):

Este segundo caso es más delicado pues:

Puede ocurrir que la verdadera regresión de la población sea curva y por lo tanto la recta es sólo una aproximación a esta curva en el tramo ajustado, acentuándose esta curvatura fuera del rango estudiado

Archivo Trigo

En un ensayo sobre trigo se desea cuantificar la relación que hay entre la disponibilidad de nitrógeno en el suelo y la cantidad de nitrógeno en la planta

Se obtuvieron datos para 12 parcelas, en las que se registró el contenido de nitrógeno en el suelo (x) y los valores promedios de nitrógeno por planta (y).

Conclusiones generales

A partir de los resultados del análisis, pueden establecerse las siguientes conclusiones:

1) La pendiente de la recta de regresión es significativamente distinto de cero, lo cual permite afirmar que entre el contenido de N en el suelo y el contenido de N en la planta existe una **relación lineal significativa** (y positiva)

Conclusiones generales

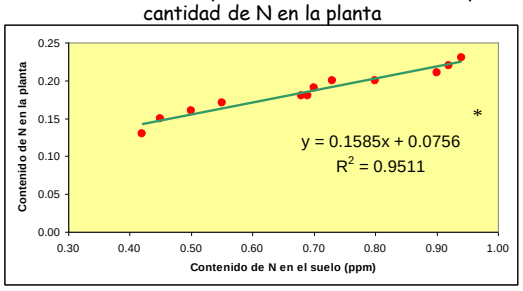
A partir de los resultados del análisis, pueden establecerse las siguientes conclusiones:

1) La pendiente de la recta de regresión es significativamente distinto de cero, lo cual permite afirmar que entre el contenido de N en el suelo y el contenido de N en la planta existe una **relación lineal significativa** (y positiva)

2) El origen de la recta de regresión es significativamente distinto de cero, pero generalmente **contrastar la hipótesis $\alpha = 0$ carece de utilidad**, pues no contiene información sobre la relación lineal entre "x" e "y"

Conclusiones generales

Relación entre la disponibilidad de N en el suelo y la cantidad de N en la planta



* Regresión estadísticamente significativa al 5 % (evaluada con el estadístico F)

Conclusiones generales

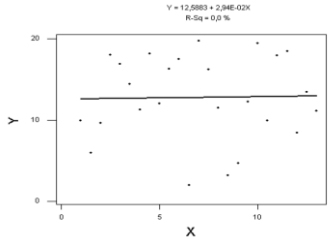
Si rechazamos HO:

Concluir que un modelo es significativo no implica necesariamente que es el mejor modelo, ya que puede que de acuerdo al diagrama de dispersión la tendencia sea curvilinea……

Conclusiones generales

Si Aceptamos HO: existen dos posibilidades:

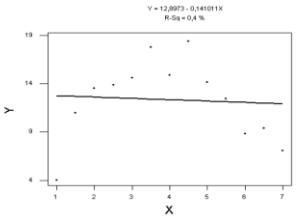
1) La variable "x" no sirve para explicar linealmente el comportamiento de las "y"



Conclusiones generales

Si Aceptamos HO: existen dos posibilidades:

2) La variable "x" contribuye a explicar la variación de "y", pero en forma no lineal



En tal caso, se debería investigar ciertos ajustes curvilíneos, sugeridos por el diagrama de dispersión

¿Por qué elegir una regresión lineal?

1) Son fáciles de manejar
2) Muchas veces son buenas aproximaciones a funciones más complicadas
3) La recta puede resultar ser una buena aproximación a una curva en un rango restringido de ella
4) La relación entre las variables puede ser realmente lineal

Supuestos Modelo de regresión lineal simple

$$Y = \alpha + \beta x + \varepsilon$$

- Y: variable dependiente o respuesta,
- X: variable independiente,
- α y β : parámetros del modelo,
- ε: error aleatorio o variación aleatoria (no explicada por el modelo) *(se supone NID(0,σ^2)).*

Supuestos Modelo de regresión lineal simple

ε: se supone NID(0,σ^2)

-Los errores de dos observaciones son independientes (Supuesto independencia)

-Los errores presentan una distribución normal (se distribuyen en forma de campana de Gauss) (Supuesto normalidad)

-La media de los errores es cero

-Su varianza es la misma cualquiera sea el valor de x (Supuesto de homocedasticidad u homogeneidad de varianzas)

Supuestos Modelo de regresión lineal simple

La verificación de los supuestos se realiza en la práctica a través de los predictores de los términos de error aleatorio que son los residuos aleatorios asociados a cada observación

Por lo tanto los supuestos pueden verificarse, mediante el análisis de los **RESIDUOS**.

Residuos

e_{ij} = VALOR OBSERVADO – VALOR PREDICHO

(eij = residuo asociado a una unidad experimental)

Se calcula como la diferencia entre el valor observado de la variable respuesta y el valor esperado estimado (predicho) bajo el modelo lineal especificado.

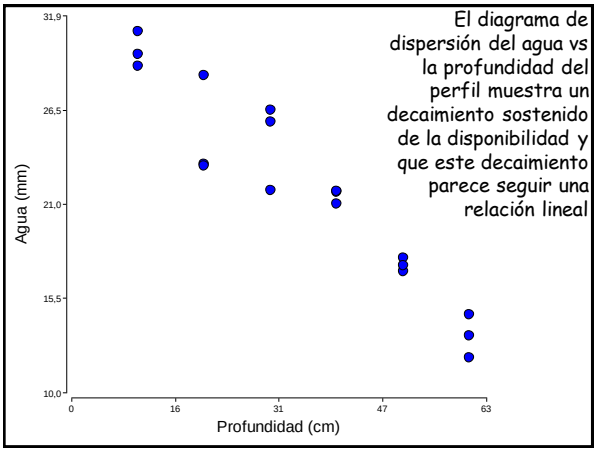
Ejemplo (Archivo Agua (en Infostat))

Lámina de agua en los perfiles del suelo de un cultivo

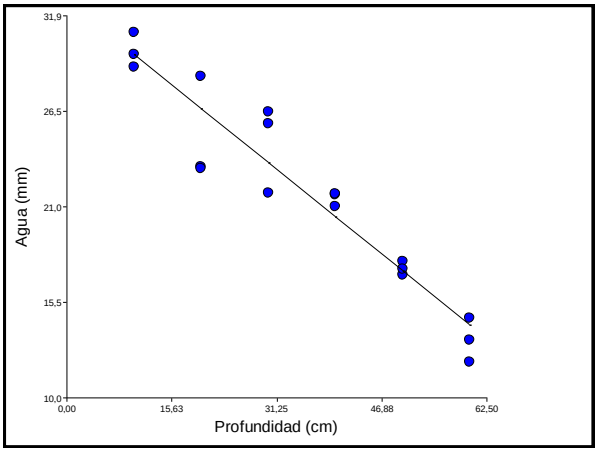
El archivo [Agua] contiene datos de disponibilidad de agua en un cultivo de soja en los distintos perfiles del suelo hasta una profundidad de 60 cm, obtenidos a los 100 días desde la emergencia. La disponibilidad de agua se expresa en milímetro de lámina de agua. Los valores de profundidad corresponden a 10, 20, 30, 40, 50 y 60 cm, pero el contenido de agua corresponde a los perfiles que van de [0-10) cm, [10-20) cm, etc. El propósito de este estudio es cuantificar cómo cambia la disponibilidad de agua con la profundidad del perfil analizado en un cultivo de soja. Los datos son parte de un estudio es más ambicioso que pretende comparar el efecto de distintos cultivares sobre el perfil de agua en el suelo. En esta aplicación sólo consideramos un cultivar. Para cada perfil hay tres repeticiones correspondientes a tres puntos de muestreo dentro de la parcela experimental.

Variable dependiente (y): agua

Variable independiente (x): profundidad



El diagrama de dispersión del agua vs la profundidad del perfil muestra un decaimiento sostenido de la disponibilidad y que este decaimiento parece seguir una relación lineal



Análisis de regresión lineal

Variable	N	R^2	R^2 Aj	ECMP	AIC	BIC
Agua (mm)	18	0,90	0,90	4,18	77,04	79,71

Coeficientes de regresión y estadísticos asociados

Coef	Est.	E.E.	LI(95%)	LS(95%)	T	p-valor	CpMallows	VIF
const	32,83	0,99	30,72	34,93	33,08	<0,0001		
Profundidad (cm)	-0,31	0,03	-0,37	-0,26	-12,20	<0,0001	141,25	1,00