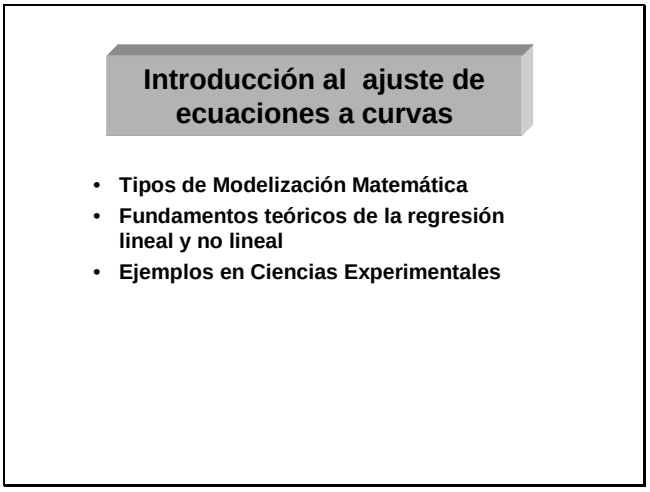
Ajuste de ecuaciones a curvas: introducción a la regresión lineal y no lineal

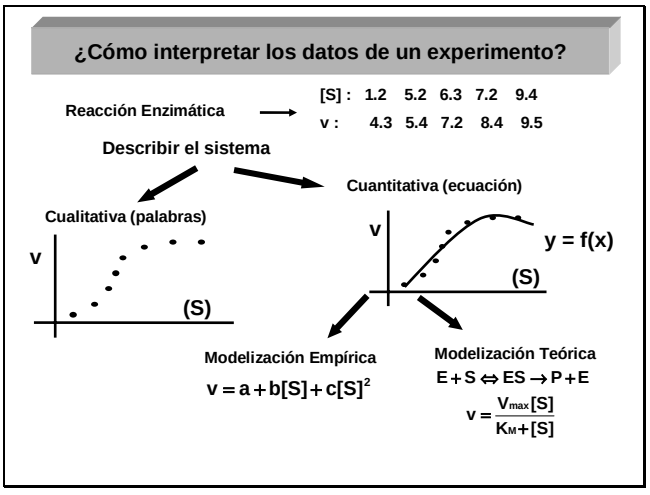
(F.J. Burguillo, Facultad de Farmacia, Universidad de Salamanca)



Diapositiva 1

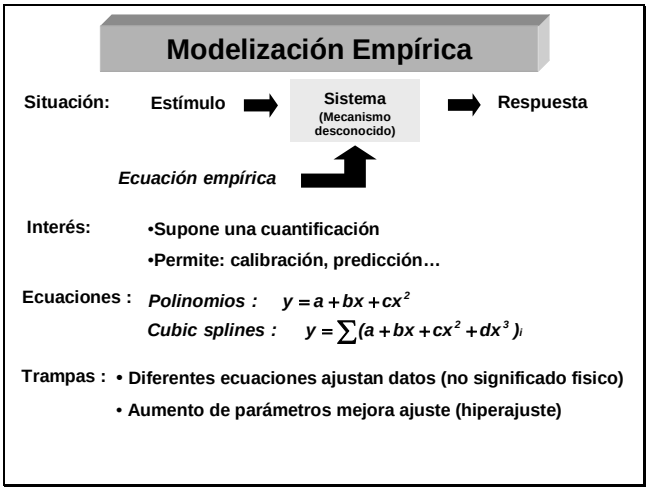
1) Si hay alguna técnica que con mayor frecuencia utilizan los investigadores ésta es la ***regresión***, incluso muchos de ellos usan casi exclusivamente la ***regresión no lineal***. Conviene pues estudiar los fundamentos de esta técnica.

En esta sesión se analizarán los distintos tipos de modelos matemáticos que se suelen emplear en el ajuste de curvas, se revisarán los fundamentos de la regresión lineal y no lineal por mínimos cuadrados y se pondrán varios ejemplos sobre el ajuste de modelos a datos experimentales.



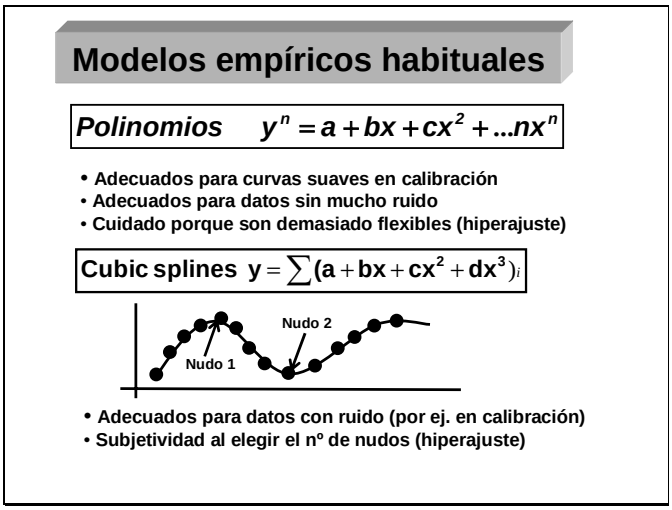
Diapositiva 2

2) El ajuste de curvas surge cuando el investigador trata de interpretar los datos de un experimento. Pensemos por ejemplo en una reacción enzimática (diapositiva 2). Los resultados se describen mejor cuando se encuentra una ecuación que se ajusta a los datos. Ese es el objetivo de la ***Modelización Matemática***: obtener ecuaciones que describan el comportamiento de los sistemas. Estas ecuaciones pueden ser de dos tipos: empíricas (*Modelización Empírica*) o deducidas en base a una teoría física (*Modelización Teórica*).



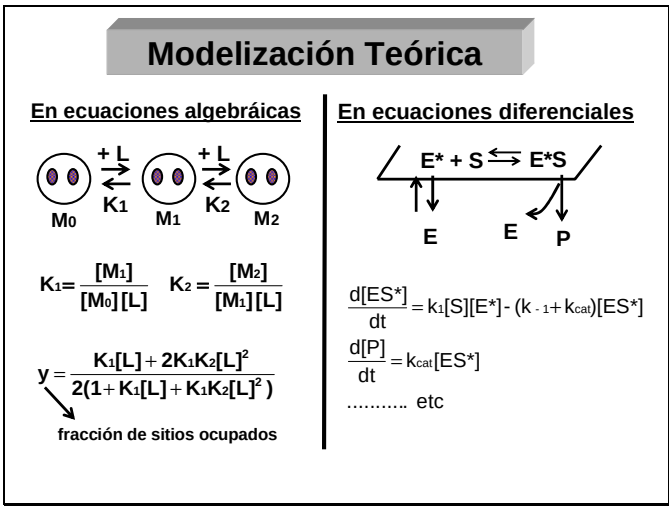
Diapositiva 3

3) La ***Modelización Empírica*** trata de encuentrar una ecuación cualquiera que cierre con los datos del sistema, independientemente de que esa ecuación tenga o no significado físico sobre lo que está ocurriendo en el sistema. Supone ya una cierta cuantificación y permite aspectos operacionales como la calibración, predicción y simulación. Por otra parte, unos mismos datos se pueden interpretar igualmente bien con diferentes ecuaciones, pero conviene elegir siempre aquellas que tenga menor número de parámetros.



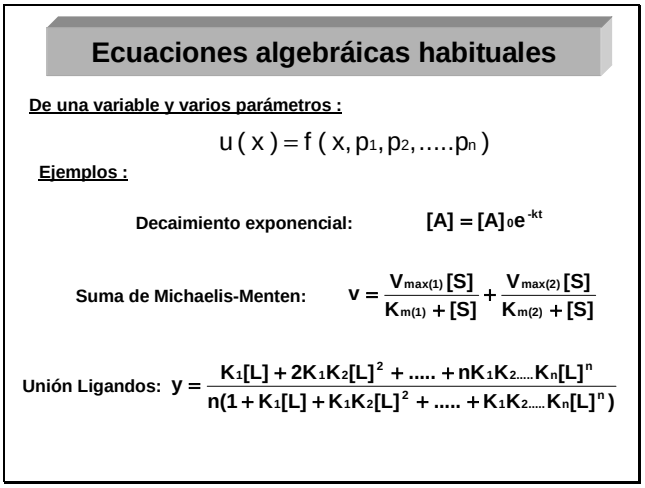
Diapositiva 4

4) Los modelos empíricos más habituales son los polinomios de distinto grado y los tramos de cúbicas (cubil splines). Algunas de sus ventajas e inconvenientes aparecen recogidos en la diapositiva 4.



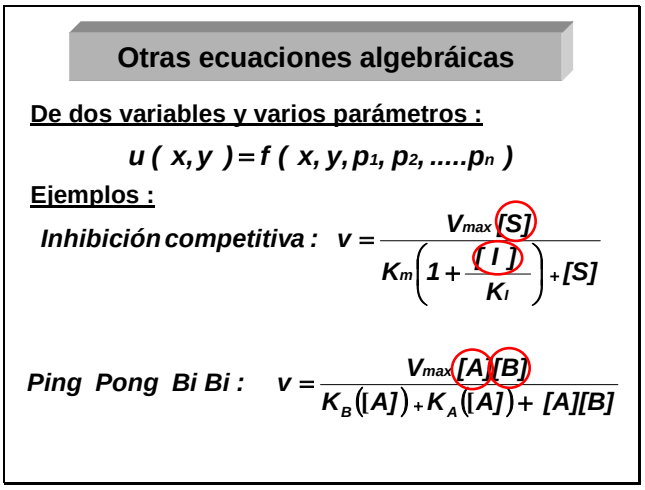
Diapositiva 5

5) La ***Modelización Teórica*** se hace normalmente en base a dos estrategias: modelos en ecuaciones algebráicas para los sistemas estáticos y modelos en ecuaciones diferenciales para los dinámicos. Primero abordaremos los modelos en ecuaciones algebráicas, que es el caso más sencillo, y más tarde los modelos en ecuaciones diferenciales. En la diapositiva 5 puede verse un ejemplo de cada tipo.



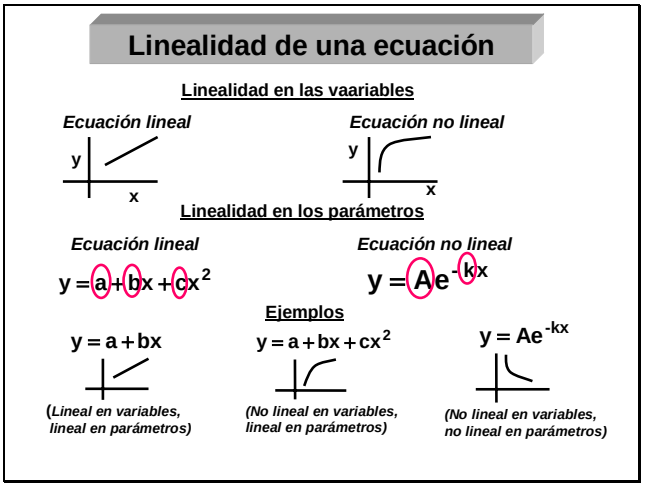
Diapositiva 6

6) En cuanto a las ***ecuaciones algebráicas*** habituales en Ciencias Experimentales, éstas son normalmente de una variable independiente y uno o varios parámetros. Algunos ejemplos de esas ecuaciones aparecen en la diapositiva 6.



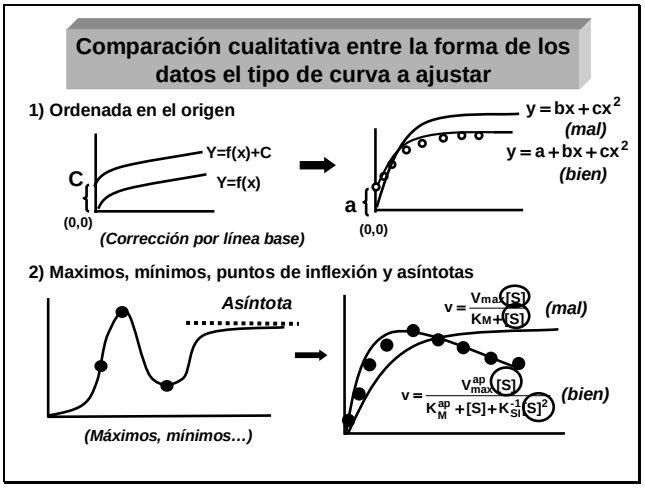
Diapositiva 7

7) También suelen darse en Ciencia las ecuaciones algebráicas de dos variables y varios parámetros. Algunos de estas ecuaciones son clásicas en Bioquímica y se muestran en la diapositiva 7. Por ejemplo, en la Inhibición Competitiva, las dos variables independientes serían [S] y [I] y los parámetros serían V_{max}, K_m y K_I.



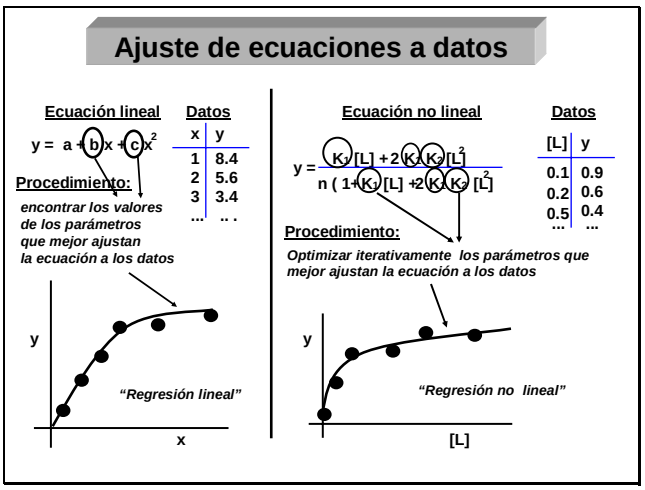
Diapositiva 8

8) En el caso de una ecuación algebráica con una variable independiente y otra dependiente, los conceptos de linealidad y no linealidad de la ecuación se pueden referir bien a las variables o a los parámetros. Una ecuación se dice que es ***lineal en las variables*** cuando su representación "y" *frente a* "x" es una recta y ***lineal en los parámetros*** cuando, considerada la x como una constante, la dependencia de y con los parámetros es combinación de sumas y restas. Los respectivos conceptos de ***no lineal*** es justo lo contrario de lo anteriormente expuesto (ver diapositiva 8).



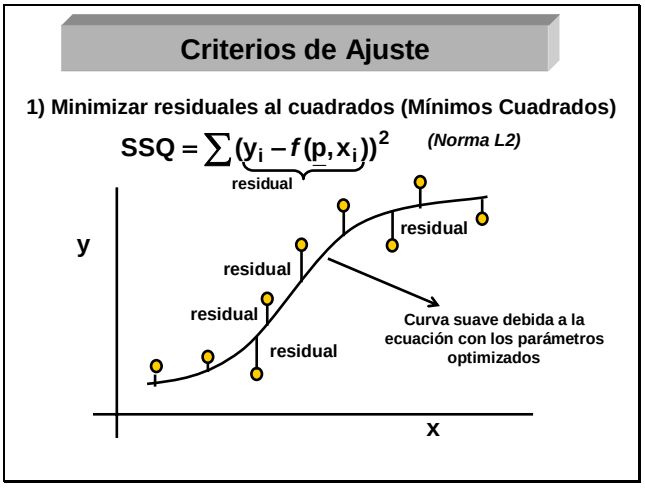
Diapositiva 9

9) Un aspecto a tener en cuenta a la hora de elegir una ecuación como modelo, es comprobar que el tipo de curva que predice nuestra ***ecuación concuerda con el comportamiento cualitativo de los datos experimentales***: ¿Pasa la curva de la ecuación por el origen ó está desplazada un cierto factor constante?, ¿Esa curva es monótona creciente o decreciente?, ¿Puede tener un máximo, un mínimo o un punto de inflexión?, ¿La curva tiende a cero, tiende a algún otro tipo de asíntota?, ¿Cierran todas esas singularidades de la curva predicha por nuestra ecuación con la tendencia de los datos?.



Diapositiva 10

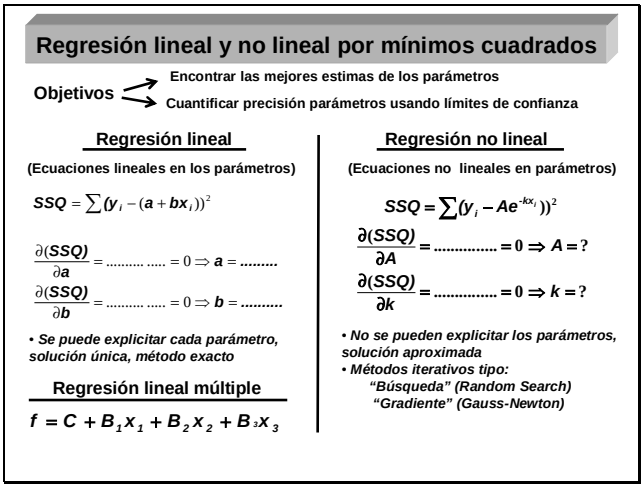
10) El paso siguiente sería el ajuste de la ecuación elegida a los datos experimentales. Este procedimiento consiste en encontrar los valores de los parámetros que mejor ajustan la ecuación a los datos (Diap. 10). Estrictamente hablando, debiera decirse ***"ajuste de la ecuación a los datos"*** y no "ajuste de los datos a la ecuación", ya que lo que se trata de "amoldar" (ajustar) es la ecuación (moviendo los valores de los parámetros) y no los datos, que son invariables. Con este sentido también se habla de "ajuste de curvas" a datos (curve fitting).



Diapositiva 11

11) Como definición de ajuste se utiliza normalmente el ***criterio de los mínimos cuadrados***, que consiste en obtener aquellos valores de los parámetros que minimizan el sumatorio de residuales al cuadrado (ver diapositiva 11). Siendo los residuales las distancias verticales de los puntos a la curva de ajuste.

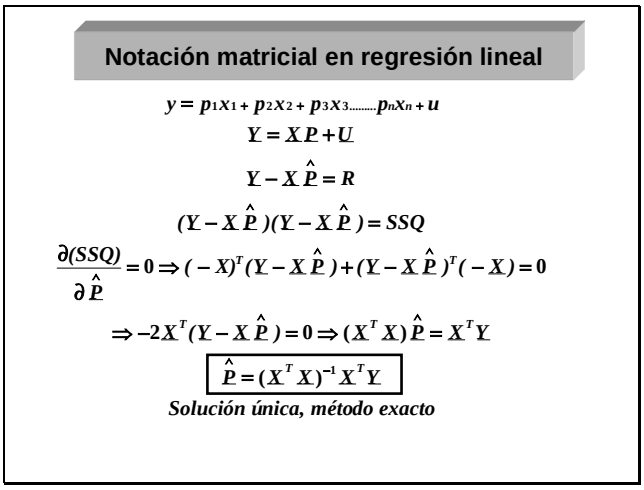
Este criterio es muy sensible a los datos atípicos, por lo que se han desarrollado otros criterios más "robustos": a) minimizar las distancias verticales absolutas y b) minimizar la distancia absoluta más grande. Pero estos criterios son menos utilizados.



Diapositiva 12

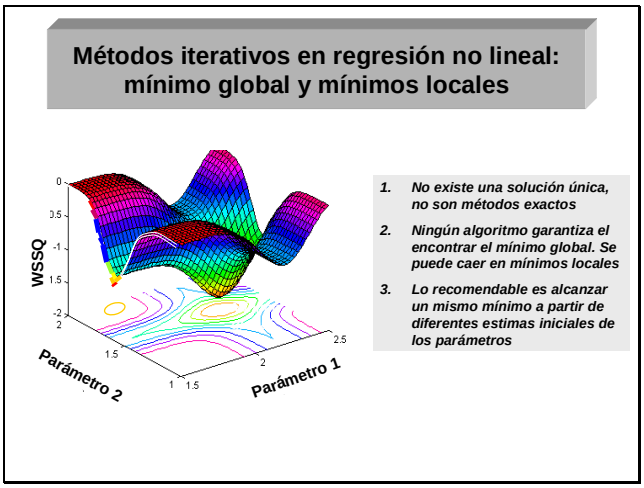
12) Se denomina *Regresión* al proceso general de ajustar una ecuación a unos datos. Y cuando esto se hace mediante el criterio de los mínimos cuadrados se habla de ***regresión lineal y no lineal por mínimos cuadrados***, según que la ecuación a ajustar sea lineal o no lineal en los parámetros (Diap. 12). En ambos casos el objetivo es el mismo: encontrar las mejores estimas de los parámetros y cuantificar la precisión de los mismos.

En el caso de la regresión lineal la solución es única y el método es exacto, mientras que en la regresión no lineal la solución es aproximada y el método es iterativo (de búsqueda, de gradiente, ...).



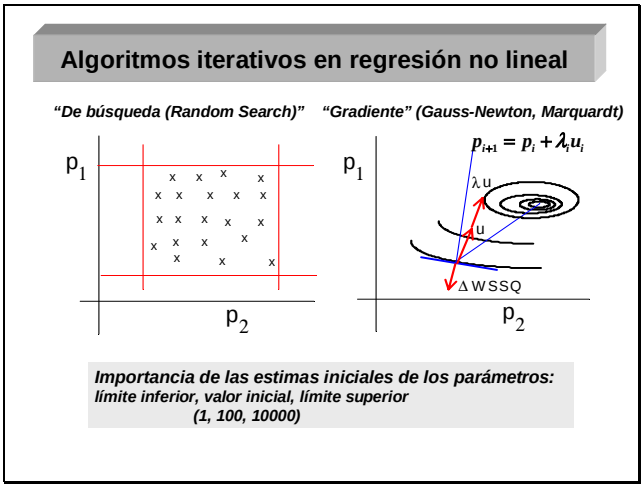
Diapositiva 13

13) El procedimiento matemático de la regresión lineal es directo. Se basa en aplicar la condición de mínimo, es decir que la derivada del sumatorio de residuales al cuadrado (SSQ) respecto a cada parámetro ha de valer cero, lo que permite despejar el valor de cada parámetro y obtener un valor único y exacto. Cuando hay varias variables y varios parámetros, el problema se maneja mejor en notación matricial, cuyo desarrollo se muestra en la diapositiva 13.



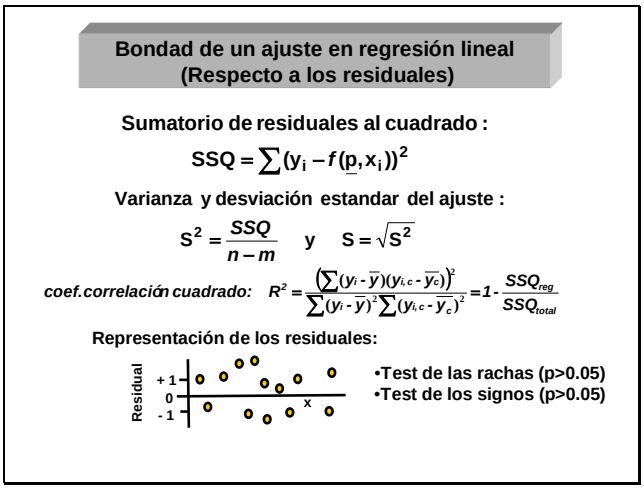
Diapositiva 14

14) En la regresión no lineal el problema de encontrar los parámetros óptimos ya no tiene una solución única de tipo explícito. En su lugar, hay que utilizar métodos iterativos, que tratan de buscar con diferentes estrategias el mínimo de SSQ. Un problema adicional es que ninguno de estos métodos garantiza el que se haya encontrado el mínimo global, existiendo la posibilidad de que se haya "caido" en un mínimo local (diap. 14). La única alternativa es probar con distintas estimas iniciales de los parámetros y ver que se llega a un mismo mínimo que podemos presumir como global.



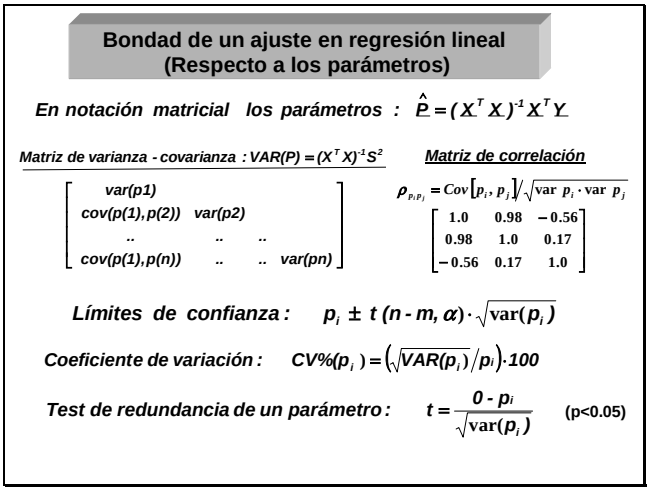
Diapositiva 15

15) Acabamos de ver que la regresión no lineal opera siempre con métodos iterativos a la hora de encontrar el mínimo del sumatorio de residuales (SSQ). Estos métodos son de búsqueda directa y de gradiente. Entre los primeros destaca el método de búsqueda al azar (random search), que consiste en ir probando en los distintos puntos de una rejilla de los parámetros hasta encontrar el mejor. Por su parte los métodos de gradiente se basan en las derivadas de SSQ respecto a los parámetros, y las diferencias entre ellos radica en la forma en que calculan la dirección de búsqueda "**u**" y la longitud del paso "**λ**" (ver Diap. 15).



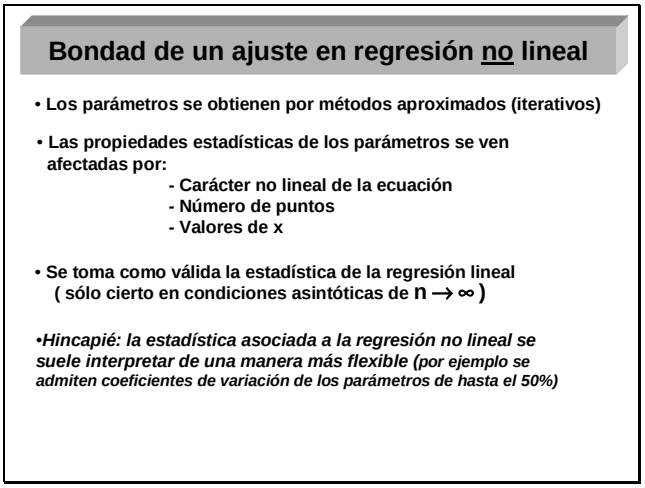
Diapositiva 16

16) Para discernir acerca de la ***bondad de un ajuste*** se utilizan diferentes criterios en la ***regresión lineal***. Unos se refieren a los ***residuales***: como son el valor del sumatorio de residuales al cuadrado, la varianza y la desviación estándar del ajuste, el coeficiente de correlación al cuadrado, la distribución gráfica de los residuales (al azar, con rachas), el test estadístico de las rachas, el test de los signos... etc (ver Diap. 16).



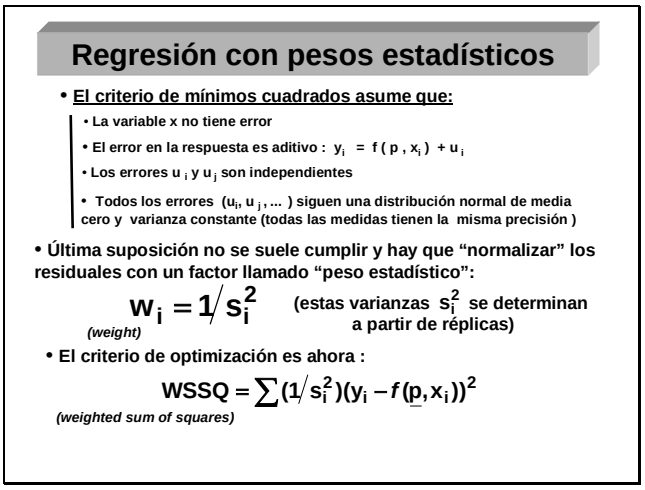
Diapositiva 17

17) Otros criterios de ***bondad de un ajuste*** se refieren a los ***parámetros***: como son las varianzas de los parámetros (dadas por la matriz de varianza-covarianza) y las correlaciones de los parámetros (matriz de correlación), los límites de confianza de los parámetros, los coeficientes de variación de los parámetros, el test de redundancia de un parámetro (su valor es tan próximo a cero que puede despreciarse)....etc (ver diapositiva 17).



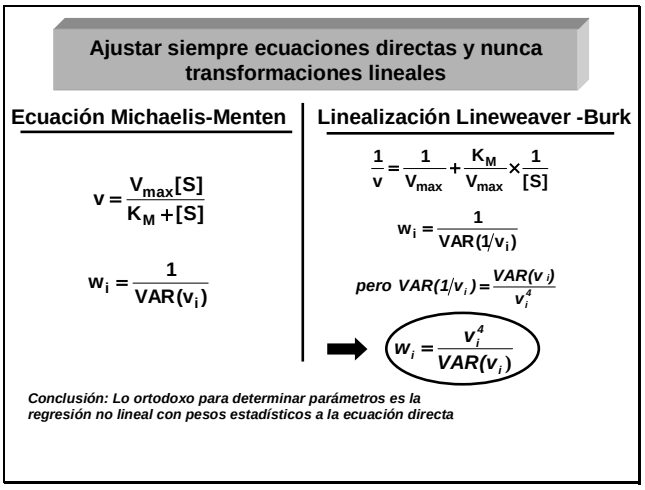
Diapositiva 18

18) En la regresión no lineal, la estadística asociada no es exacta y, por defecto, se acepta como aproximada la estadística de la regresión lineal expuesta más arriba, lo cual solo sería cierto en condiciones asintóticas de infinito número de puntos. En este sentido, se suele ser más flexible a la hora de interpretar los indicadores acerca de la bondad del ajuste. Así, por ejemplo, en regresión no lineal se suelen admitir coeficientes de variación de los parámetros de hasta un 50%.



Diapositiva 19

19) El criterio de los mínimos cuadrados asume que el error al medir la variable dependiente es aditivo, que son independientes unos errores de otros y que en conjunto siguen una distribución de media cero y varianza constante. A este tipo de regresión se la denomina ***regresión sin pesos estadísticos***. Cuando esta suposición no es cierta (que es la mayoría de las veces), se hace necesario dar más "importancia" (más peso) a los datos de menor error, frente a los de mayor error (menos peso). Para ello se corrigen los residuales con un factor llamado ***peso estadístico*** que se define como el ***inverso de la varianza*** (Diap.13) y que viene a ser un factor de normalización de residuales muy dispares. Estos valores de varianza se suelen obtener a partir de réplicas de cada dato experimental.



Diapositiva 20

20) Para ajustar ecuaciones no lineales en los parámetros, se han utilizado mucho las ***transformaciones lineales*** (dobles inversos, logaritmos..), seguidas de regresión lineal sin pesos estadísticos (calculadoras). Esta práctica es desaconsejable, ya que no considera la propagación del error en la ecuación transformada, por lo que la estima de los parámetros y sus límites de confianza son erróneos. Para ajustar ecuaciones no lineales, lo más correcto es la regresión no lineal con pesos estadísticos a la ecuación directa (y=f(x)), sin transformación alguna (ver diapositiva 20).

Discriminación entre modelos

Lo habitual es que se dude entre modelos alternativos dentro de una secuencia, por ejemplo en una mezcla de isoenzimas :

$$v = \frac{V_{max(1)}[S]}{K_{M(1)}+[S]} + \frac{V_{max(2)}[S]}{K_{M(2)}+[S]} + \ldots\ldots + \frac{V_{max(n)}[S]}{K_{M(n)}+[S]}$$

1) Conviene comparar la bondad de los 2 ajustes rivales:
WSSQ, residuales, test de las rachas,
límites de confianza de los parámetros…etc

2) Se debe aplicar el test "F" (modelos jerarquizados) :

$$F = \frac{[(SSQ_1 - SSQ_2)/(m_2 - m_1)]}{SSQ_2/(n - m_2)}$$

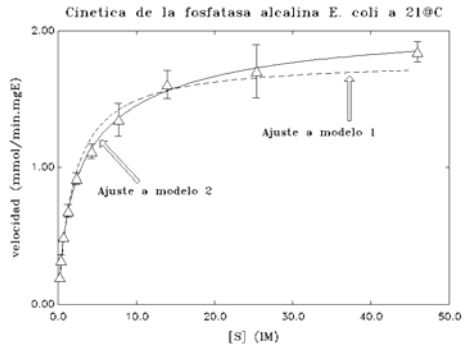
Si F > F(95%) se acepta modelo 2

Si F < F(95%) se acepta modelo 1

3) Otros criterios para modelos jerarquizados y no jerarquizados son:
Criterio AIC de Akaike, Mallows Cp

Diapositiva 21

Discriminación por superposición de ajustes



Diapositiva 22

Ajuste a ecuaciones de 2 variables

Ecuación:

$$\textit{Inhibición competitiva :}\quad v = \frac{V_{max}[S]}{K_m\left(1 + \frac{[I]}{K_I}\right) + [S]}$$

Datos:

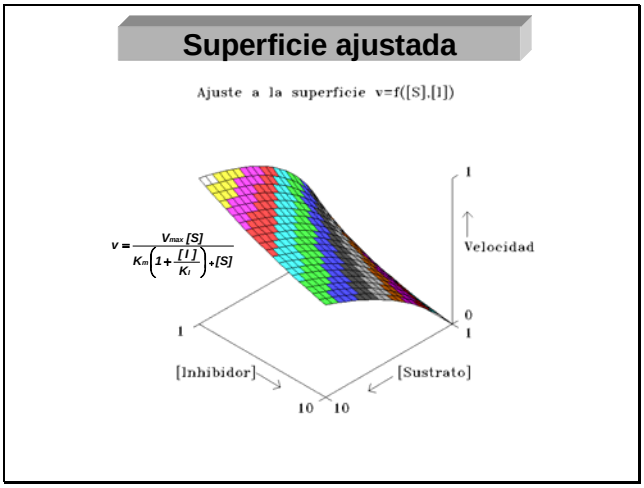
Inhibidor :	**1**	**1**	**1**	**1**	**2**	**2**	**2**	**2**	.....
Sustrato :	**2**	**4**	**6**	**8**	**2**	**4**	**6**	**8**	.....
velocidad :	**5.2**	**6.3**	**7.1**	**9.1**	**3.2**	**5.2**	**6.4**	**7.5**	.....

Diapositiva 23

21) Cuando se analiza un sistema bioquímico, lo normal es que se dude entre modelos alternativos dentro de una secuencia jerárquica (1 isoenzima, 2 isoenzimas...), se impone pues alguna estrategia para ***discriminar entre modelos rivales***. Esta discriminación suele hacerse comparando la bondad de los distintos ajustes y en base al ***test estadístico "F"***, que valora si es o no estadísticamente significativa la mejora que experimenta habitualmente el sumatorio de residuales al cuadrado al pasar de una ecuación de menos parámetros a otra de más parámetros (Diapositiva 21).

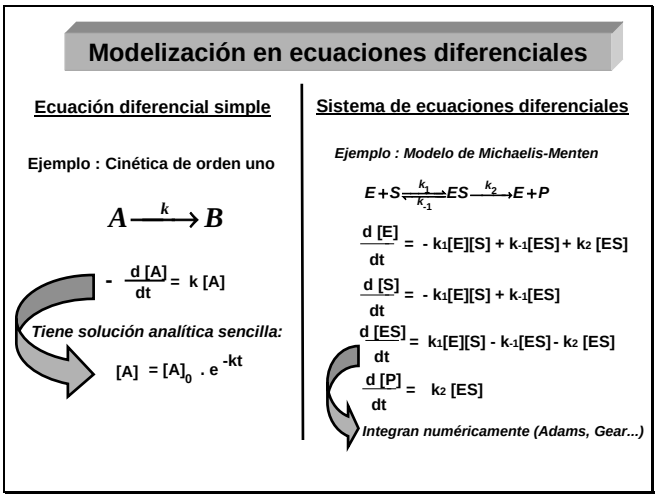
22) Otro criterio para la discriminación entre modelos es la ***superposición de los ajustes respectivos***, con el fin de observar si los puntos se distribuyen al azar a ambos lados de la curva ajustada (buen ajuste) o si presentan tendencia a las rachas (mal ajuste). Esta discriminación visual se puede hacer tanto en la representación directa de la ecuación (diapositiva 22), como en otro tipo de transformaciones de la misma, principalmente transformaciones lineales que son muy intuitivas. Pero esta estrategia de transformaciones lineales sólo es válida a efectos gráficos de discriminación o de presentación final de resultados, ya que el ajuste y la determinación de los parámetros se deben hacer en el espacio directo.

23) Hasta ahora nos hemos referido al caso de sólo una variable independiente. Pero ocurre en Ciencia que es muy frecuente el estudio de sistemas con 2 variables independientes o más, como ocurre por ejemplo con la inhibición competitiva en Bioquímica, en la que la velocidad de reacción depende de la concentración del sustrato y del inhibidor (Diap. 23).



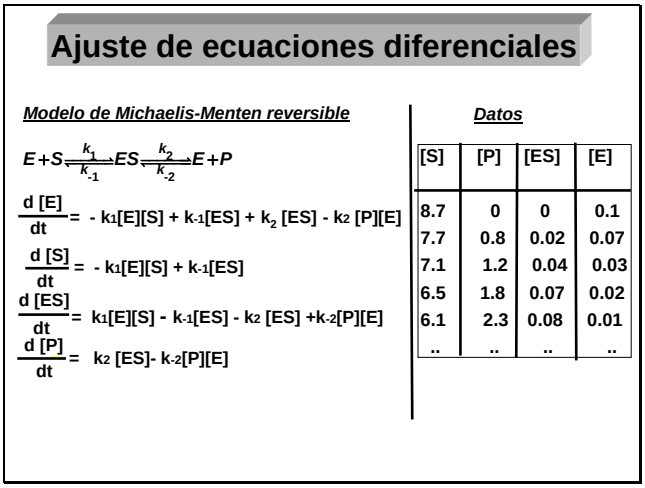
Diapositiva 24

24) Estas ecuaciones en dos variables independientes también se pueden ajustar por técnicas de regresión no lineal. En este caso lo que se ajusta no es una curva sino una superficie, como puede observarse en la diapositiva 24 para la inhibición competitiva, donde se ha representado la velocidad en el eje "z", la concentración de sustrato en el eje "x" y la del inhibidor en el eje "y".



Diapositiva 25

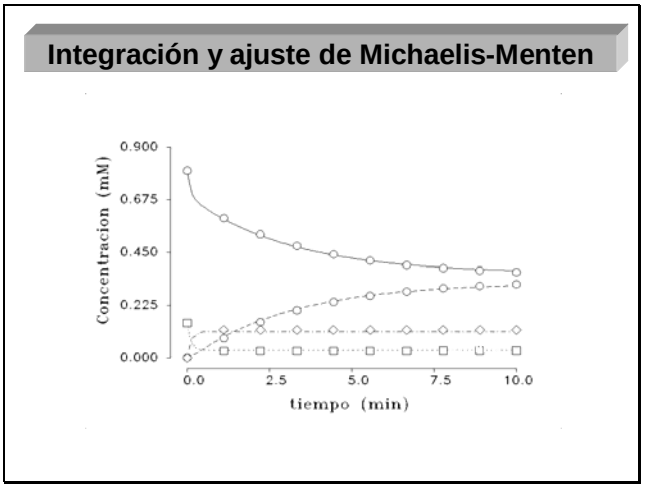
25) La modelización en ***ecuaciones diferenciales*** puede presentar diferentes formas. Si se trata de una ***ecuación diferencial simple***, lo usual es que sea de una variable independiente, una variable dependiente y varios parámetros. Este caso se suele integrar la ecuación diferencial analíticamente y realizar el ajuste en base a la ecuación integrada correspondiente. Cuando se trata de varias ecuaciones simultáneas, el caso mas frecuente es el de un sistema de ***ecuaciones diferenciales ordinarias***, en el que solo hay una variable independiente (normalmente el tiempo), varias variables dependientes y diferentes parámetros (Diap. 25). Su ajuste a los datos experimentales se aborda por técnicas de integración numérica y optimización. Los sistemas de ecuaciones con más de una variable independiente (por ejemplo tiempo y distancia), llamados en ***ecuaciones diferenciales en derivadas parciales***, son menos frecuentes y más difíciles de tratar.



Diapositiva 26

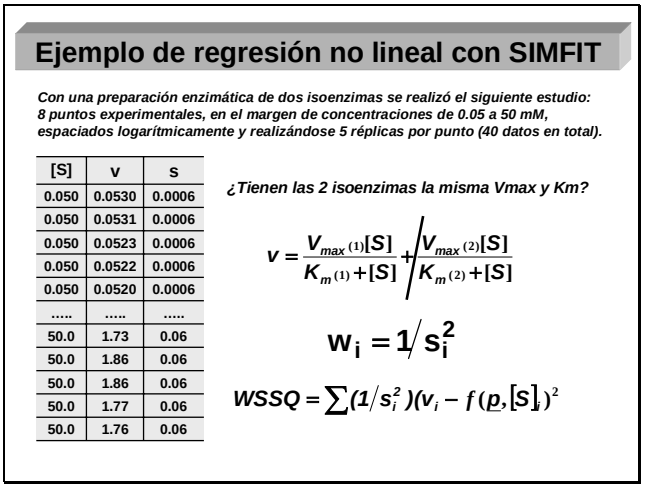
26) Un ejemplo de ecuaciones diferenciales ordinarias para el modelo de "Michaelis-Menten reversible" se muestra en la diapositiva 26.

Si quisiéramos ajustar ese sistema de ecuaciones a los datos empíricos de todas o algunas de las variables, necesitaríamos un programa que hiciera simultáneamente una integración numérica de las ecuaciones diferenciales seguida de una comparación con los puntos experimentales. Y así iterativamente hasta que el ajuste alcanzara la convergencia.



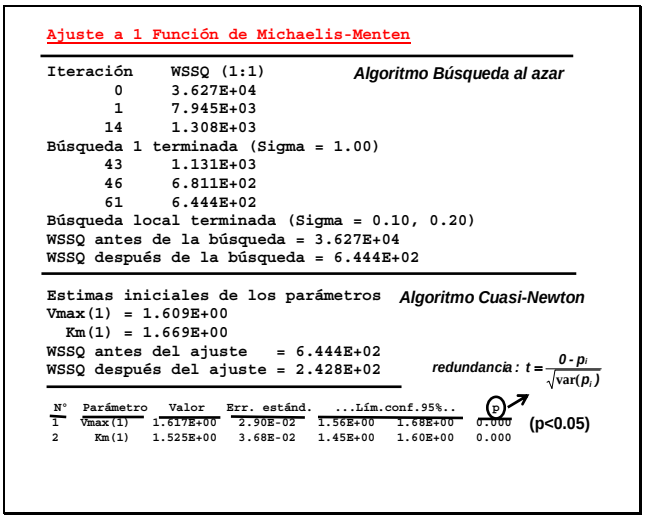
Diapositiva 2

27) En la diapositiva 27 puede observarse el ajuste de las ecuaciones diferenciales comentadas en el punto anterior a unos datos experimentales simulados. Estas integraciones y ajustes con ecuaciones diferenciales ordinarias se pueden hacer en SIMFIT con el programa DEQSOL.



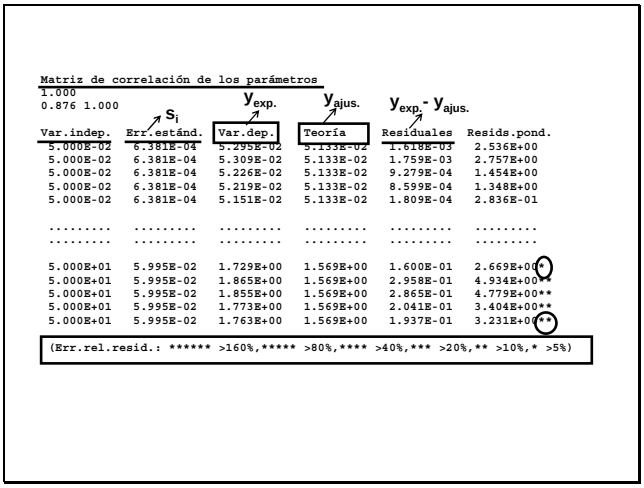
Diapositiva 28

28) A modo de caso práctico veamos como abordar con SIMFIT algún ajuste por regresión no lineal. Imaginemos una preparación enzimática de dos isoenzimas, con la que se hace un estudio cinético con el objetivo de ver si las 2 isoenzimas tienen la misma V_{max} y K_m y en su caso determinar estos valores. En esencia se trata de discriminar si la ecuación de velocidad requiere 1 o 2 términos de Michaelis-Menten (ver Diap. 28), para lo cual vamos a hacer un ajuste de regresión no lineal con pesos estadísticos a los dos modelos alternativos.



Diapositiva 29

29) Primero el programa ajusta a los datos la función con sólo 1 término de Michaelis-Menten. Comienza con un algoritmo de búsqueda al azar que va probando diferentes valores de los parámetros y se va quedando con los que dan un menor valor de WSSQ. Esos valores entran como estimas iniciales a un algoritmo de gradiente Cuasi-Newton, que sigue optimizando el WSSQ hasta que se alcanza la convergencia. Hecho esto el programa muestra los valores de los parámetros, su error estándar, sus límites de confianza y el valor de "p" de redundancia de cada parámetro (Diap. 29).



Matriz de correlación de los parámetros
1.000
0.876 1.000

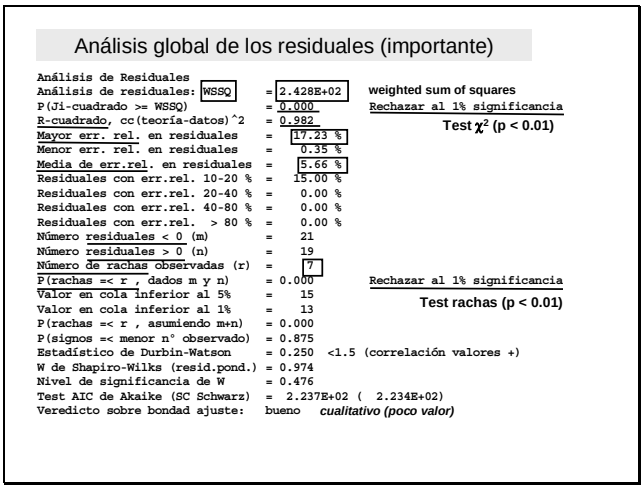
s_i $y_{exp.}$ $y_{ajus.}$ $y_{exp.} - y_{ajus.}$

Var.indep.	Err.estánd.	Var.dep.	Teoría	Residuales	Resids.pond.
5.000E-02	6.381E-04	5.295E-02	5.133E-02	1.618E-03	2.536E+00
5.000E-02	6.381E-04	5.309E-02	5.133E-02	1.759E-03	2.757E+00
5.000E-02	6.381E-04	5.226E-02	5.133E-02	9.279E-04	1.454E+00
5.000E-02	6.381E-04	5.219E-02	5.133E-02	8.599E-04	1.348E+00
5.000E-02	6.381E-04	5.151E-02	5.133E-02	1.809E-04	2.836E-01
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
5.000E+01	5.995E-02	1.729E+00	1.569E+00	1.600E-01	2.669E+00*
5.000E+01	5.995E-02	1.865E+00	1.569E+00	2.958E-01	4.934E+00*
5.000E+01	5.995E-02	1.855E+00	1.569E+00	2.865E-01	4.779E+00**
5.000E+01	5.995E-02	1.773E+00	1.569E+00	2.041E-01	3.404E+00**
5.000E+01	5.995E-02	1.763E+00	1.569E+00	1.937E-01	3.231E+00**

(Err.rel.resid.: ****** >160%,***** >80%,**** >40%,*** >20%,** >10%,* >5%)

30) Sigue después mostrando la matriz de correlación de los parámetros y una tabla completa con los valores de "$y_{experimental}$", "$y_{ajustada}$" y los residuales ($y_{experimental}$-$y_{ajustada}$). Si el error relativo del residual es grande, se levanta un asterisco o más de uno a la derecha del residual, indicando que el error relativo es > del 5 %, > del 10 % ....etc (ver Diap. 30).

Diapositiva 30

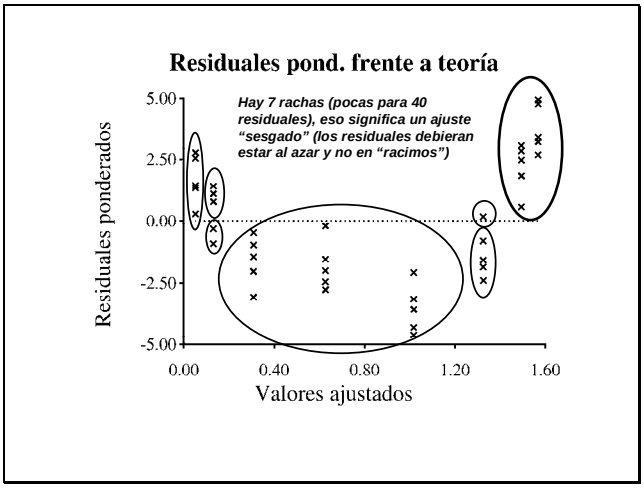


Análisis global de los residuales (importante)

Análisis de Residuales
Análisis de residuales: WSSQ = 2.428E+02 weighted sum of squares
P(Ji-cuadrado >= WSSQ) = 0.000 Rechazar al 1% significancia
R-cuadrado, cc(teoría-datos)^2 = 0.982 Test χ^2 (p < 0.01)
Mayor err. rel. en residuales = 17.23 %
Menor err. rel. en residuales = 0.35 %
Media de err.rel. en residuales = 5.66 %
Residuales con err.rel. 10-20 % = 15.00 %
Residuales con err.rel. 20-40 % = 0.00 %
Residuales con err.rel. 40-80 % = 0.00 %
Residuales con err.rel. > 80 % = 0.00 %
Número residuales < 0 (m) = 21
Número residuales > 0 (n) = 19
Número de rachas observadas (r) = 7
P(rachas =< r , dados m y n) = 0.000 Rechazar al 1% significancia
Valor en cola inferior al 5% = 15 Test rachas (p < 0.01)
Valor en cola inferior al 1% = 13
P(rachas =< r , asumiendo m+n) = 0.000
P(signos =< menor nº observado) = 0.875
Estadístico de Durbin-Watson = 0.250 <1.5 (correlación valores +)
W de Shapiro-Wilks (resid.pond.) = 0.974
Nivel de significancia de W = 0.476
Test AIC de Akaike (SC Schwarz) = 2.237E+02 (2.234E+02)
Veredicto sobre bondad ajuste: bueno *cualitativo (poco valor)*

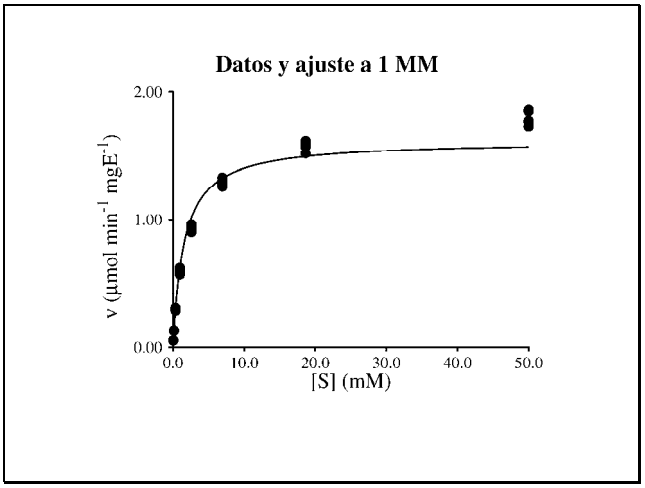
31) A continuación se muestra un análisis global de los residuales. Se da el valor de WSSQ, que si el ajuste es bueno debiera seguir una distribución Ji-cuadrado con n-par. grados de libertad, extremo que se encarga de analizar la "p" correspondiente, que en este caso vale 0.000 (p < 0.05) y hace que se levante una "bandera" a la derecha con un "Rechazar al 1 % de significancia". Luego sigue el valor de R^2, la media de los errores relativos de los residuales, el nº de residuales positivos y negativos, el nº de rachas, la "p" del test de las rachas, que en este caso vale 0.000 (p < 0.05) y hace que se levante la "bandera" correspondiente. (Ver Diap. 31).

Diapositiva 31



32) El programa permite hacer también una gráfica de los residuales, en este caso de residuales ponderados (considerando sus pesos) en ordenadas frente a los valores teóricos ajustados (Diap. 32). Para este ajuste se observan 7 rachas, que son pocas para 40 residuales, donde cabría esperar una distribución más al azar y por tanto mayor número de rachas. Esto significa un ajuste sesgado, lo que unido al análisis hecho en el punto anterior (31), nos va llevando a la conclusión de que la bondad de este ajuste es escasa.

Diapositiva 32



Diapositiva 33

33) La mejor confirmación de que el ajuste no es bueno, como venimos apuntando, es representar los puntos experimentales junto a la función ajustada (Diap. 33). A la vista de esta gráfica se comprueba que el ajuste está sesgado, ya que los últimos puntos los ha dejado claramente por encima de la curva.

Ajuste a 2 Michaelis-Menten

Iteración	WSSQ (2:2)
0	3.627E+04
1	1.045E+04
7	3.393E+03
21	1.262E+03
30	8.976E+02
143	5.505E+02

Algoritmo búsqueda al azar

Búsqueda 1 terminada (Sigma = 1.00)

185	5.462E+02
195	4.145E+02
202	3.354E+02
222	2.044E+02

Búsqueda local terminada (Sigma = 0.10, 0.20)

Para la búsqueda al azar 2:2
N° de mejoras en 320 ciclos = 9
WSSQ antes de la búsqueda = 3.627E+04
WSSQ después de la búsqueda = 2.044E+02

Algoritmo Cuasi-Newton

Estimas iniciales de los parámetros
Vmax(1) = 1.530E+00
Vmax(2) = 6.222E-01
Km(1) = 1.500E+00
Km(2) = 1.091E+02
WSSQ antes del ajuste = 2.044E+02
WSSQ después del ajuste = 3.442E+01

Ajuste 2:2 Función de Michaelis-Menten

N°	Parámetro	Valor	Err. estánd.	...Lím.conf.95%..		p
1	Vmax(1)	9.317E-01	6.70E-02	7.96E-01	1.07E+00	0.000
2	Vmax(2)	1.033E+00	8.81E-02	8.55E-01	1.21E+00	0.000
3	Km(1)	9.823E+00	2.39E+00	4.97E+00	1.47E+01	0.000
4	Km(2)	1.033E+00	6.43E-02	9.03E-01	1.16E+00	0.000

Las 4 "p" son < 0.05 , parámetros distintos "0"

Diapositiva 34

34) Automáticamente, el programa comienza a ajustar la función con 2 términos de Michaelis-Menten (Diap. 34). Entra primero el algoritmo de búsqueda y después el de gradiente Cuasi-Newton, mostrando finalmente la tabla con los 4 parámetros de esta función (2 V_{max} y 2 K_m), así como sus límites de confianza y sus valores de "p" de redundancia del parámetro. Las "p" son todas < 0.05 y los límites de confianza parecen razonables. Luego la bondad de los parámetros parece adecuada.

Matriz de correlación de los parámetros

1.000
-0.834 1.000
0.990-0.869 1.000
0.930-0.593 0.882 1.000

Var.indep.	Err.estánd.	Var.dep.	Teoría	Residuales	Resids.pond.
5.000E-02	6.381E-04	5.295E-02	5.242E-02	5.310E-04	8.322E-01
5.000E-02	6.381E-04	5.309E-02	5.242E-02	6.720E-04	1.053E+00
5.000E-02	6.381E-04	5.226E-02	5.242E-02	-1.590E-04	-2.491E-01
5.000E-02	6.381E-04	5.219E-02	5.242E-02	-2.270E-04	-3.557E-01
5.000E-02	6.381E-04	5.151E-02	5.242E-02	-9.060E-04	-1.420E+00
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
5.000E+01	5.995E-02	1.729E+00	1.791E+00	-6.235E-02	-1.040E+00
5.000E+01	5.995E-02	1.865E+00	1.791E+00	7.345E-02	1.225E+00
5.000E+01	5.995E-02	1.855E+00	1.791E+00	6.415E-02	1.070E+00
5.000E+01	5.995E-02	1.773E+00	1.791E+00	-1.825E-02	-3.044E-01
5.000E+01	5.995E-02	1.763E+00	1.791E+00	-2.865E-02	-4.779E-01

(Err.rel.resid.: ****** >160%,***** >80%,**** >40%,*** >20%,** >10%,* >5%)

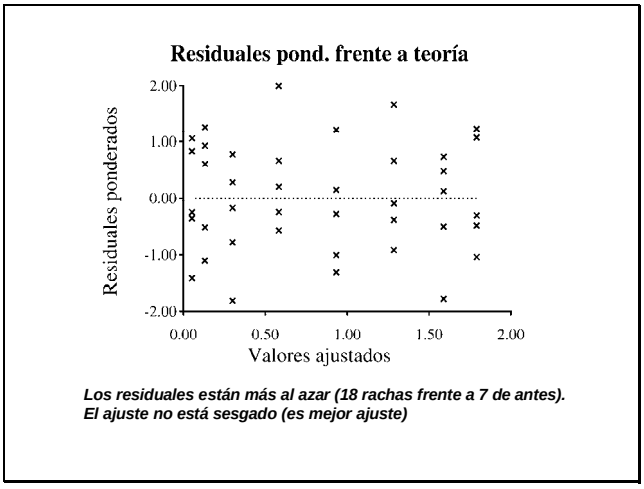
Diapositiva 35

35) A continuación se muestra para este ajuste la matriz de correlación de los parámetros y la tabla con los valores de Var. dep. ($y_{exp.}$), Teoría ("$y_{ajust.}$)" y los residuales ($y_{exp.}$-$y_{ajus.}$). En esta ocasión los residuales parecen pequeños, ya que no se levantan asteriscos a la derecha de los mismos, indicando que el error relativo no es > del 5 %, > del 10 % ....etc (ver diapositiva 30).

Análisis global de los residuales para 2 MM

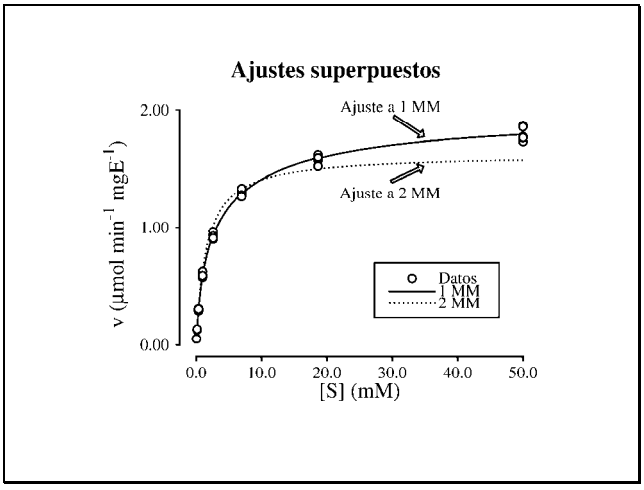
```
Análisis de Residuales
Análisis de residuales:  WSSQ          = 3.442E+01   (disminuyó (antes 2.43E+02))
P(Ji-cuadrado >= WSSQ)                 = 0.544       → Test χ² (buen ajuste p > 0.05)
R-cuadrado, cc(teoría-datos)^2         = 0.998
Mayor err. rel. en residuales          =    6.64 %
Menor err. rel. en residuales          =    0.21 %
Media de err.rel. en residuales        =    1.96 %   (disminuyó (antes 5.66 %))
Residuales con err.rel. 10-20 %        =    0.00 %
Residuales con err.rel. 20-40 %        =    0.00 %
Residuales con err.rel. 40-80 %        =    0.00 %
Residuales con err.rel.  > 80 %        =    0.00 %
Número residuales < 0 (m)              =    21
Número residuales > 0 (n)              =    19
Número de rachas observadas (r)        =    18       (aumentó (antes 7 ))
P(rachas =< r , dados m y n)           =  0.217      (test rachas (buen ajuste ( p > 0.05 ))
Valor en cola inferior al 5%           =    15
Valor en cola inferior al 1%           =    13
P(rachas =< r , asumiendo m+n)         = 0.261
P(signos =< menor n° observado)        = 0.875
Estadístico de Durbin-Watson           = 2.019
W de Shapiro-Wilks (resid.pond.)       = 0.982
Nivel de significancia de W            = 0.776
Test AIC de Akaike (SC Schwarz)        =  1.495E+02 (  1.489E+02)
Veredicto sobre bondad ajuste:         increible
```

Diapositiva 36



Los residuales están más al azar (18 rachas frente a 7 de antes). El ajuste no está sesgado (es mejor ajuste)

Diapositiva 37

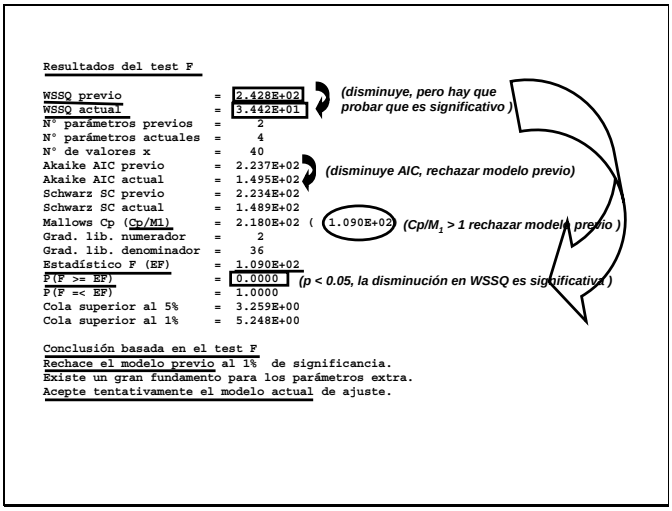


Diapositiva 38

36) A continuación se muestra el análisis de los residuales de este ajuste. El valor de WSSQ ha disminuido apreciablemente respecto al del modelo previo. La "p" del test ji-cuadrado para WSSQ vale ahora 0.544 ($p > 0.05$) lo que indica un buen ajuste. El valor de R^2 vale ahora 0.998 (frete al 0.982 del modelo previo). La media de los errores relativos de los residuales ha bajado de 5.66 % en el modelo previo ha 1.96 % del actual. El nº de rachas ha subido de 7 en el modelo anterior a 18 en el actual, con un "p" en el test de las rachas de 0.217 ($p > 0.05$), lo que significa una distribución más al azar de los residuales, mejor ajuste (Diap. 36).

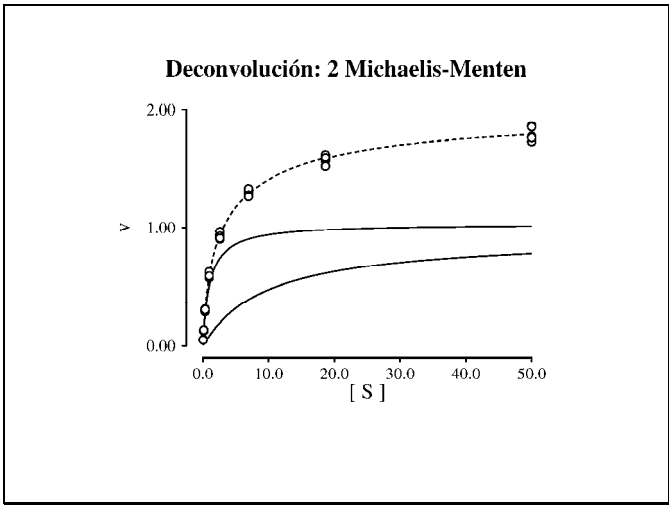
37) En este ajuste (diapositiva 37), la representación de los residuales ponderados (con pesos) frente a los valores teóricos ajustados presenta una distribución mucho más al azar que en el ajusto al modelo previo. Esto significa un ajuste mucho menos sesgado, lo que unido al análisis hecho en el punto anterior (36), nos va llevando a la conclusión de que la bondad del ajuste actual es mayor que la del ajuste previo.

38) Otra confirmación de que el ajuste actual a 2 MM es mejor que el ajuste previo a 1 MM, nos lo brinda la superposición de los dos ajustes a los puntos experimentales (Diap. 38). A la vista de esta gráfica se comprueba que el ajuste a 2 MM se adapta mejor a los puntos que el ajuste a 1 MM. Esta confirmación visual es muy valiosa para un investigador, pero parece que hace falta algún criterio estadístico que lo sancione cuantitativamente. Entro estos criterios se encuentra el test F, el criterio de Akaike y otros.



Diapositiva 39

39) El propio programa se encarga de aplicar el test "F" a los dos WSSQ obtenidos, el del modelo previo (242.8) y el del actual (34.42). No hay duda de que el WSSQ se ha reducido en ocho veces, pero también se ha incrementado el nº de parámetros de 2 a 4. ¿Es significativa esta disminución del WSSQ?. De eso se encarga el test "F", proporcionándonos en esta caso una p=0.0000 (p < 0.01), por lo que podemos rechazar el modelo previo al 1 % de significancia y quedarnos con el modelo actual, porque aunque tiene más parámetros éstos estarían justificados.



Diapositiva 40

40) Una última prueba, acerca de si 2 términos de MM estarían justificados para ajustar los datos experimentales, sería el hacer una deconvolución de la función a los dos términos que la forman y comprobar si su participación es suficientemente relevante. En este caso (ver Diap. 40) se puede apreciar que los dos términos contribuyen casi por igual a la función total. Parece razonable suponer que cada una de las 2 isoenzimas tiene una cinética diferente, pero globalmente se solapan para dar una única curva v-[S] que es la que se observa experimentalmente.

Bibliografía

1) William G. Bardsley, ***"SIMFIT: Reference manual"*** (2004), http://www. simfit. man. ac.uk.
2) H. J. Motulsky and A. Christopoulos, ***"Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting"*** (2003), http://www. graphpad.com.
3) Leah Edelstein-Keshet, ***"Mathematical Models in Biology"*** (1988) , McGraw-Hill.
4) Paul Doucet and Peter B. Sloep, ***"Mathematical Modeling in the Life Sciences"*** (1992) , Ellis Horword.
5) Laszlo Endrenyi (Ed.), **"Kinetic Data Analysis. Design and Analysis of Enzyme and Pharmacokinetic Experiments"** (1981), Plenum Press.