Unidad 5 (Extensión)

Métodos cuantitativos de análisis – Significación de Parámetros de un ajuste

- ✓ **Método de cuadrados mínimos en el caso de datos con errores**
- ✓ **Incertidumbre en la extracción de parámetros de un ajuste**
- ✓ **Ensayo de significación estadística de parámetros de un ajuste**
- ✓ **Bandas de predicción de valores**
- ✓ **Normalidad de una distribución estadística**
- ✓ **Caso de datos con error en las dos variables**

Método de cuadrados mínimos –Datos con errores - Regresión lineal -

Consideremos el caso de un conjunto de mediciones (X_i, Y_i), con error en el valor de Y_i dado por σ_i. cuyas representaciones gráficas se muestran en la Figura 1. El objetivo de esta sección es describir el procedimiento estadístico que permite obtener la línea que mejor ajusta los datos experimentales, línea de regresión, y las incertezas asociadas en su determinación. Asimismo se desea tener un modo de estimar las incertezas asociadas a la estimación de un dado valor Y_0, a partir de un valor X_0.

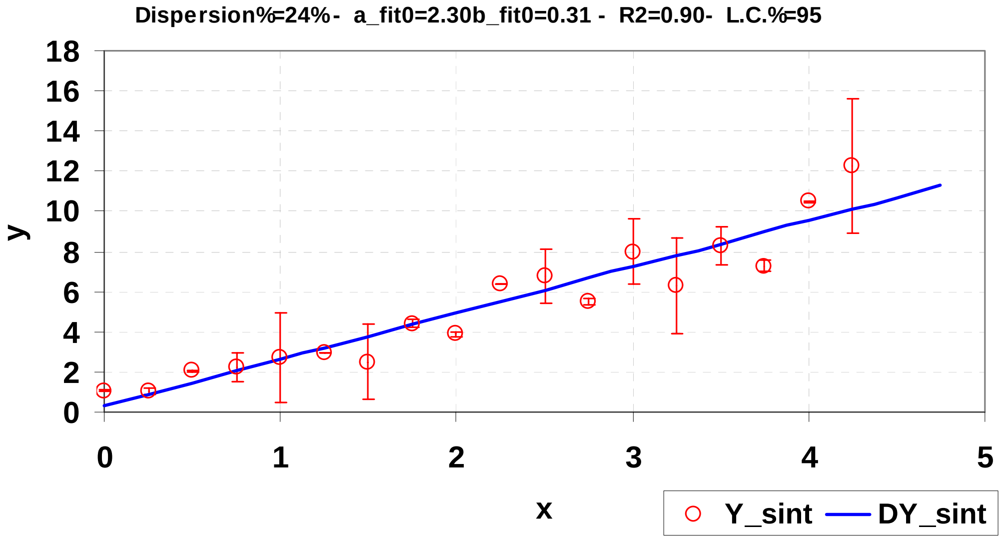

Figura 1.- Representación gráfica de un conjunto de datos experimentales (X_i, Y_i) con errores en el eje Y dado por los valores σ_i . La línea continua azul es la recta obtenida por cuadrados mínimos.

Incertidumbre en la extracción de parámetros de un ajuste

La recta que mejor ajusta los datos viene dad por la ecuación:

$$Y(x) = a \cdot x + b \tag{1}$$

Definimos el valor de Chi-Cuadrado, χ^2, como:

$$\chi^2 = \sum_{i=1}^{N} w_i \cdot (Y_i - a \cdot x_i - b)^2 \tag{2}$$

Aquí W_i es un factor de peso o ponderación que se puede definir de distintos modos según el problema en estudio. Un modo usual de pesar los datos es hacerlo usando sus respectivos errores σ_i del siguiente modo:

$$w_i = \frac{1}{\sigma_i^2}\,. \tag{3}$$

con

$$W = \sum_i w_i \tag{4}$$

Si todos los datos tienen igual ponderación, es decir si, w_i=1, entonces $W=N$ = *número total de datos*. Desde luego, la expresión (3) representa solo una de las tantas formas en que pueden ponderarse los datos. La elección más adecuada de los factores de ponderación depende del problema especifico en consideración.

El método de cuadrados mínimos consiste en elegir como los mejores valores de a y b aquellos valores que minimicen el valor de χ^2 ec.(2). El resultado de este procedimiento resulta en[1-4]:

$$a = \frac{W \cdot \left(\sum YX\right) - \left(\sum X\right) \cdot \left(\sum Y\right)}{W \cdot \left(\sum X^2\right) - \cdot \left(\sum X\right)^2} = \frac{<YX> - <X> \cdot <Y>}{\left(<X^2> - <X>^2\right)} = \frac{Cov(YX)}{S_X^2} \tag{5}$$

$$b = \frac{\left(\sum Y\right) \cdot \left(\sum X^2\right) - \left(\sum X\right) \cdot \left(\sum Y \cdot X\right)}{W \cdot \left(\sum X^2\right) - \cdot \left(\sum X\right)^2} = \frac{<X^2> \cdot <Y> - <X \cdot Y> \cdot <X>}{\left(<X^2> - <X>^2\right)} \tag{6}$$

o bien

$$b = <Y> - a \cdot <X> \tag{6'}$$

Donde usamos la notación:

$$\sum Y \equiv \sum_{i=1}^{N} w_i \cdot Y_i\,, \tag{7}$$

$$\sum X^n \equiv \sum_{i=1}^{N} w_i \cdot X_i^n\,, \tag{8}$$

$$\sum Y \cdot X \equiv \sum_{i=1}^{N} w_i \cdot Y_i \cdot X_i \tag{9}$$

y así sucesivamente.

También definimos los valores medios de *y* y *x* como:

$$\bar{Y} \equiv < Y > \equiv \sum_{i=1}^{N} w_i \cdot Y_i \Big/ W \,, \tag{10}$$

$$\bar{X} \equiv < X > \equiv \sum_{i=1}^{N} w_i \cdot X_i \Big/ W \,, \tag{11}$$

$$< X^n > \equiv \sum_{i=1}^{N} w_i \cdot X_i^n \Big/ W \,, \tag{12}$$

Las desviaciones estándar vienen dadas por:

$$S_x^2 \equiv \sum_{i=1}^{N} w_i \cdot \left(X_i - \bar{X}\right)^2 \Big/ W = \left(\frac{N-1}{N}\right) \cdot Var(X) = < X^2 > - < X >^2 \,, \tag{13}$$

y

$$S_Y^2 \equiv \sum_{i=1}^{N} w_i \cdot \left(Y_i - \bar{Y}\right)^2 \Big/ W = \left(\frac{N-1}{N}\right) \cdot Var(Y) = < Y^2 > - < Y >^2 \,, \tag{14}$$

Los coeficientes de correlación se definen en forma similar:

$$S_{YX} \equiv Cov(Y, X) = \sum_{i=1}^{N} w_i \cdot \left(X_i - \bar{X}\right) \cdot \left(Y_i - \bar{Y}\right) \Big/ W = < YX > - < Y >< X > \,, \tag{15}$$

y

$$R \equiv \frac{< YX > - < Y >< X >}{S_X \cdot S_Y} = \frac{Cov(YX)}{S_X \cdot S_Y} \,. \tag{16}$$

El error típico de estimación de *Y* sobre *X*, está relacionado con el valor de Chi-cuadrado, $\chi_N^{\ 2}$, por:

$$\chi_N^2 = \frac{\sum w_i \cdot (Y_i - Y(X_i))^2}{W} = Error.típico(YX)^2 \tag{17}$$

También se define el valor de Chi-cuadrado por grados de libertad: χ^2_v:

$$\chi_v^2 = sd^2 = \frac{N}{(N-2)} \cdot \frac{\sum w_i \cdot (Y_i - Y(X_i))^2}{W} = \frac{N}{N-2} \cdot \chi_N^2 \tag{18}$$

El parámetro, $\chi_N^{\ 2}$ se relaciona con la covarianza de *XY*, para el caso de una ajuste lineal (5)y (6), por la relación:

$$\chi_N^2 = S_Y^2 - \frac{\left[< (X-\overline{X})\cdot(Y-\overline{Y})>\right]^2}{S_Y^2} = S_Y^2 - \left[\frac{Cov(Y,X)}{S_X}\right]^2 \tag{19}$$

La variación total, S_t=S_Y, da una medida de cómo los puntos Y_i se distribuyen alrededor del valor medio de Y. S_t se define como:

$$N\cdot S_t^2 = \sum_{i=1}^{N} w_i\cdot\left(Y_i-\overline{Y}\right)^2 = N\cdot S_Y^2 \tag{20}$$

La variación explicada, S_{inex}, mide la calidad del modelo, $Y(X_i)$, para explicar los datos observados, Y_i. Este nombre surge del hecho que $\varepsilon_i=(Y_i-Y(X_i))$ tienen una distribución estadística al azar. S_{inex} se define como

$$\sum_{i=1}^{N} w_i\cdot\left(Y_i-Y(X_i)\right)^2 = W\cdot S_{inex}^2 = W\cdot\chi_N^2 = \frac{N-2}{N}\cdot W\cdot sd^2 \tag{21}$$

La variación explicada S_{exp}, se define por:

$$\sum_{i=1}^{N} w_i\cdot\left(Y(X_i)-\overline{Y}\right)^2 = N\cdot S_{\exp}^2 \tag{22}$$

A partir de (20), (21) y (22) se demuestra que:

$$S_t^2 = S_{\exp}^2 + S_{inex}^2 \tag{23}$$

de donde tenemos la propiedad:

$$R^2 = \frac{S_{\exp l}^2}{S_t^2} = 1-\frac{S_{inex}^2}{S_t^2} = 1-\frac{\chi_N^2}{S_Y^2} \le 1 \tag{24}$$

También tenemos que:

$$Cov(X,Y) = \left(\sum W_i\cdot(X_i-\overline{X})\cdot(Y_i-\overline{Y})\right)\Big/W = R\cdot S_X\cdot S_Y \tag{25}$$

Una propiedad importante de los estimadores a y b, es que si los errores o residuos de las estimaciones, ε_i:

$$\varepsilon_i^2 = \left(Y_i - Y(X_i)\right)^2, \tag{26}$$

tiene una distribución normal, si tomamos muestras sucesivas de la población, o realizamos conjuntos independientes de mediciones, cada una tendrá un valor de a y b distintos en general. Los valores a y b tendrán a su vez una distribución estadística y sus valores medios vendrán dados por $<a>=a_0$ y $<b>=b_0$ y sus desviaciones estándar, o *errores estándar*, dadas por Δa_0 y Δb_0 respectivamente por:

$$\Delta a_0 = \frac{\sqrt{\chi_N^2}}{\sqrt{N-2}\cdot S_X} = \frac{sd}{\sqrt{N-2}\cdot S_X} = \frac{a_0}{\sqrt{N-2}}\cdot\sqrt{\left(\frac{1}{R^2}-1\right)} \tag{27}$$

y

$$\Delta b_0 = \frac{\sqrt{\chi_N^2}}{\sqrt{N-2}\cdot S_X}\cdot\sqrt{<X^2>} = \frac{sd}{\sqrt{N-2}\cdot S_X}\cdot\sqrt{<X^2>} = \Delta a_0\cdot\sqrt{<X^2>}\,. \quad (28)$$

Si los errores ε_i tienen una distribución normal, la variable aleatoria *t,* definida por:

$$t = \frac{(a-<a>)}{\Delta a_0} \quad (29)$$

presenta una distribución *t*-Student, con *N-2* grados de libertad .

Para calcular la incerteza en la estimación de *a* (Δa) a partir de una muestra de tamaño *N*, con un límite de confianza de *P%*, se calcula a partir de del valor t_p, que se obtiene de la distribución *t*-Student con (N-2) grados de libertad

Probabilidad_t-Student ($t<t_p$)*=P%* (30)

Si se usa Excel® Microsoft este valor de t_p se calcula usando la función DISTR.T.INV((1-P), N-2). La incerteza Δa se calcula como:

$$\Delta a(P\%) \equiv \Delta a_p = t_p\cdot\frac{sd}{\sqrt{N}\cdot S_X} = t_p\cdot\Delta a_0 = t_p\cdot a_0\cdot\sqrt{\frac{(1/R^2-1)}{(N-2)}} \quad (31)$$

El error en *b* viene dado por:

$$\Delta b(P\%) \equiv \Delta b_p = \Delta a(P\%)\cdot\sqrt{<X^2>}\,. \quad (32)$$

Ensayo de significación estadística de parámetros de un ajuste

Un ensayo usual que es necesario realizar, es el evaluar o docimar si el valor de la pendiente (a_0) obtenida de un dado experimento, es significativamente distinta de otro ensayo que arrojo como resultado $a_0=A$. En definitiva lo que deseamos evaluar es la hipótesis nula [#]H_0: $a_0=A$ frente a la hipótesis H_1: $a_0\neq A$. Suponiendo que la variable aleatoria:

$$t = \frac{(A-a_0)}{\Delta a_0} \quad (33)$$

[#] Muchas veces formulamos una hipótesis con el único objeto de rechazarla, por ejemplo si deseamos decidir si una moneda esta cargada o trucada, formulamos la hipótesis que la moneda en buena. Estas hipótesis se denominan *hipótesis nula*[8] y se designa con H_0. La máxima probabilidad con que deseamos rechazar una hipótesis cuando debió se aceptada (*Error tipo I*), se llama *nivel de significación* y se designa con α. Valores frecuentes de α son 0.005 (5%) o 0.01 (1%).

tiene un distribución *t*-Student con N-2 grados de libertad, si deseamos docimar H_0 frente a H_1 con un nivel de confianza de *P%*, evaluamos el correspondiente coeficiente de confianza t_p, dada por relación (30). De modo tal que si:

$$a - t_p \cdot \Delta a_0 < A < a - t_p \cdot \Delta a_0 \tag{34}$$

aceptamos H_0, en caso contrario, la misma debe ser desechada. En particular, si deseamos falsear (evaluar si es posible desechar o no) la hipótesis H_0: $a_0=0$, deseamos calcular la probabilidad, *P*, que el valor efectivamente encontrado de la pendiente ($a_0 \neq 0$) sea consistente con la hipótesis H_0: $a_0=0$, vine dada por:

$$P = Dist.t\left(\frac{a_0}{\Delta a_0}, N-2\right) \tag{34}$$

Donde la probabilidad se calcula usando la distribución usada es la *t*-Student con (N-2) grados de libertad. Si se usa Excel® Microsoft este valor se calcula usando la función DISTR.T ($a_0/\Delta a_0$; *N-2, 2*), este último argumento (2) esta asociado al hecho que se usan las dos colas de la distribución. Claramente, cuanto más cercana a cero sea esta probabilidad, mayor será la confianza que tendremos en que la variable independiente (X en nuestro caso) es relevante para explicar la variación de la variable dependiente Y, y mayor es nuestra confianza en que la hipótesis H_1: $a_0 \neq 0$ es la correcta. Estas ideas pueden generalizarse aún para el caso no lineal. De este modo, para evaluar si un dado parámetro *b*, es relevante o no para explicar los datos, o sea si debe o no incluirse en el modelo, un criterio es calcular su probabilidad usando la distribución DISTR.T ($b_0/\Delta b_0$; $N-\nu$, ν), siendo ν =*número grados de libertad* y verificando si su valor supera un nivel de significación α previamente establecido, por ejemplo α=5%.

Bandas de predicción de valores

Muchas veces, el objeto de un ajuste, el obtener los parámetros del modelo con el objeto de realizar perdiciones o proyecciones de una variable dependiente, Y, para nuevos valores de la variable independiente, *X*. En otras palabras, queremos realizar predicciones con nuestro modelo. Para estimar la incerteza asociada a una proyección de un nuevo valor, calculado para un valor no medido X_0, obtenido usando la recta de regresión $Y_0=a.X_0+b$, con un límite de confianza de *P%* tenemos dos casos distintos: a) estimación de la probabilidad que un valor individual de una muestra, asociada al valor de X=X0 caiga con probabilidad P% entre $Y(X_0) - \Delta Y_{estim}$ y $Y(X_0) + \Delta Y_{estim}$.

$$\begin{aligned}\Delta Y_{estim}(X_0) &= t_p \cdot \frac{sd}{\sqrt{N}} \cdot \sqrt{\left(1 + N + (X_0 - \bar{X})^2/(S_x^2)\right)} = \\ &= t_p \cdot \sqrt{\frac{\chi_N^2}{(N-2)}} \cdot \sqrt{\left(1 + N + (X_0 - \bar{X})^2/(S_x^2)\right)} = . \\ &= t_p \cdot \Delta a_0 \cdot \sqrt{S_x^2 + (X_0 - \bar{X})^2}\end{aligned} \tag{35}$$

b) estimación de la probabilidad que un valor medio de los valores una muestra, asociada al valor de X=X0 caiga con probabilidad *P%* entre $Y(X_0)$ - ΔY_{estim_media} y $Y(X_0)$ + ΔY_{estim_media}.

$$\Delta Y_{estim\_media}(X_0) = t_p \cdot \frac{sd}{\sqrt{N}} \cdot \sqrt{\left(1 + (X_0 - \bar{X})^2 / (S_x^2)\right)} =$$
$$= t_p \cdot \sqrt{\frac{\chi_N^2}{(N-2)}} \cdot \sqrt{\left(1 + (X_0 - \bar{X})^2 / (S_x^2)\right)} = . \quad (36)$$
$$= t_p \cdot \Delta a_0 \cdot \sqrt{S_x^2 + (X_0 - \bar{X})^2}$$

Es usual indicar las *bandas de los intervalos de confianza* dadas por (36) como las bandas de confianza. También se utilizan las bandas determinadas por (35) y usualmente se las designa como *bandas de predicción*.

Para cada ordenada X_0, los valores de $\Delta Y(X_0)$ definen dos curvas: $y(x)+\Delta Y(X_o)$ y $y(x)-\Delta Y(X_o)$ entre las cuales encontraremos el *P%* de los datos observados. Estas bandas definen los límites de confianza de *P%* para las predicciones de Y. Ver figura 2.

Normalidad de una distribución estadística

Normalidad de los residuos o perturbaciones: En muchos de las secciones anteriores, supusimos que los errores de las estimaciones o residuos ε_i, EC. 26, tenían una distribución normal. Para evaluar en un dado caso de interés si esta hipótesis se cumple, se pueden realizar varios "*test de normalidad*." Tal vez el más simple y directo consiste en construir un histograma de los residuos y calcular los primeros momentos de la distribución de ε_i, m_k.k=1,2,3,… Si el histograma de los residuos empíricos, tiene forma de campana parecida a la de la distribución Normal $N(m,\sigma)$, siendo $m=m_1$ y $\sigma_2=m_2-m_1^2$. Entonces tenemos una primera indicación, muy cualitativa, una distribución normal. Un criterio algo más cuantitativo consiste en construir un histograma de las probabilidades empíricas acumulas y representarlo en escala Probabilistica. En esta escala una distribución normal tendría una tendencia lineal. Figura 3.

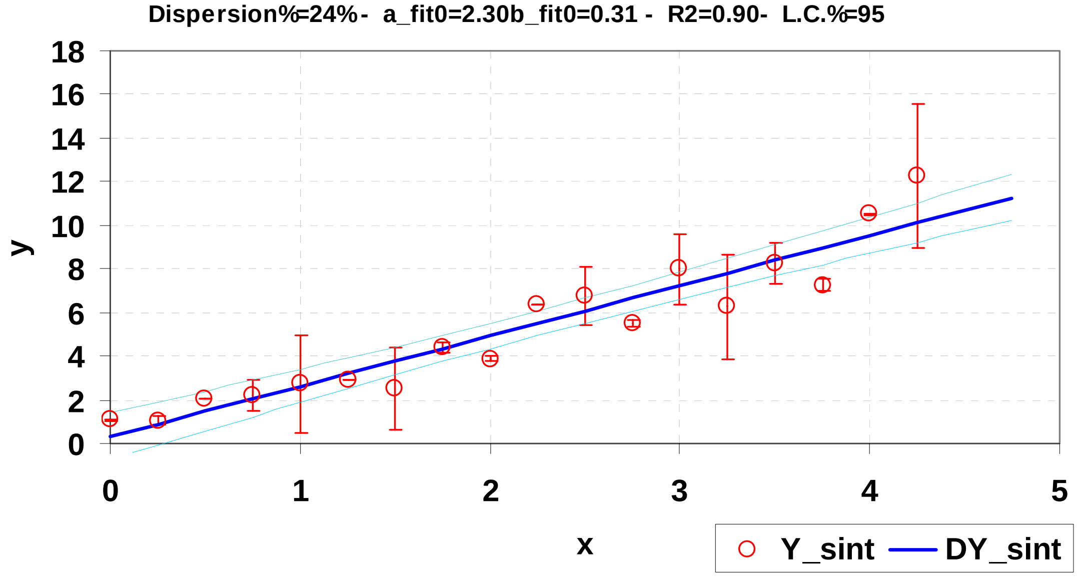

Figura 2.- Representación gráfica de un conjunto de datos experimentales (X_i, Y_i). Las bandas laterales representan los limites de confianza de *95% (bandas de confianza)*..

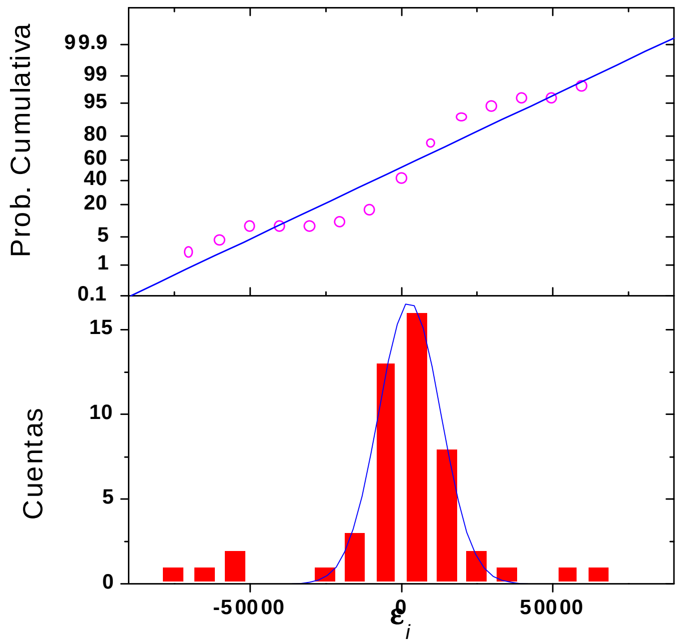

Figura 3.- Histograma de los residuos ε_i (Ec. 26). En el cuadro superior tenemos la distribución acumulada del histograma en función del los valores de ε_i. La escala vertical es Probabilistica, de modo que una distribución normal acumulada daría una línea recta en estas escala. En el cuadro inferior se muestra la distribución empírica de los residuos (barras) junto con la curva normal de igual media m y desviación estándar que la empírica.

Un criterio más cuantitativo es el de Jarque-Bera, para ello debemos calcular los coeficientes de asimetría *Asim* y la *Curtosis* (Apuntamiento) definidas por:

$$Asim = \frac{m_3}{m_2^{3/2}} \tag{37}$$

y

$$Curtosis = \frac{m_4}{m_2^2} \tag{38}$$

Para una distribución normal, la Asimetría (y todos los momentos impares) son cero y la Curtosis es igual a 3. Definimos el coeficiente *JB* como:

$$JB = N_{datos} \cdot \left[\frac{1}{6} \cdot Asim^2 + \frac{1}{24} \cdot \left(Curtosis - 2 \right)^2 \right] \Rightarrow \chi^2_{\nu=2}(JB) \tag{39}$$

Si la distribución de ε_i es normal, estos coeficientes tiene una distribución estadística asimptotica Chi-cuadrado con ν=2 grados de libertad. Es claro que para una distribución normal, el parámetro JB →0. La probabilidad, de que siendo normal la distribución de residuos empíricos, el valor de *JB* tenga un valor finito, vendrá dad por:

$$P(JB) = \int_{JB}^{\infty} \chi^2_{\nu=2}(x) \cdot dx = DISTR.CHI(JB,2) \tag{40}$$

Así si trabajamos con un límite de significación α=5%, si *JB*>6 debemos rechazar la hipótesis de que los residuos ε_i están normalmente distribuidos. En Excel, esta probabilidad se calcula con la función *DISTR.CHI(JB*,ν=2*)*.

Caso de datos con error en las dos variables

Caso de Error en ambas variables: En general las técnicas estadísticas para considerar estos casos es motivo discusión entre los distintos autores y expertos en este tema. Aquí proponemos un esquema aproximado, basado fundamentalmente en las ref.4-6.

Si las mediciones (x_i,y_i) tiene errores: Δx_i y Δy_i receptivamente. Definimos los factores de peso para cada punto como:

$$W_i = 1/\sigma_i^2 \tag{41}$$

donde:

$$\sigma_i^2 = a^2 \cdot \Delta x_i^2 + \Delta y_i^2 \,. \tag{42}$$

En general si los factores de ponderación de la variable *x* e *y* son w_{xi} y $w_{y,i}$ respectivamente, entonces:

$$w_i = \frac{w_{x,i} \cdot w_{y,i}}{a^2 \cdot w_{y,i} + w_{x,i}} \tag{43}$$

donde a es la pendiente de la recta de regresión, ec.(5). El problema es que para determinar a debemos de resolver el problema de regresión. Para ello necesitamos los factores de peso W_i, que a su vez dependen de a. Para resolver este problema podemos proceder de modo iterativo. Usamos como ponderación inicial solo los valores de $w_{yi}(=1/\Delta y_i^2)$, Con estos coeficientes, usando (5) obtenemos el valor de a, con este valor calculamos los pesos w_i usando (42) determinamos de nuevo los coeficientes w_i y a partir de la ec.(5) los nuevos coeficientes a. Iterando hasta que los sucesivos valores de a no cambien, se obtienen los parámetros de la regresión lineal buscada, o sea la regresión lineal para el caso de datos con errores en las dos variables.

Estas ideas pueden extenderse al caso no lineal, en que la función *f(x;a,b,c,...)* cuyos parámetros, *a, b, c, ...* se buscan determinar, depende de un modo no lineal de *x*. En este caso la generalización de (42) conduce al concepto de error efectivo[6]:

$$\sigma_i^2 = \left(\frac{df}{dx}\right)^2 \cdot \Delta x_i^2 + \Delta y_i^2 \,. \tag{44}$$

Análogamente, la expresión (42) para los pesos o ponderación se puede generalizar en:

$$w_i = \frac{w_{x,i} \cdot w_{y,i}}{(df/dx)^2 \cdot w_{y,i} + w_{x,i}} \,. \tag{45}$$

Bibliografía

1. P. Bevington and D. K. Robinson, *Data reduction and error analysis for the physical sciences*, 2nd ed. (McGraw Hill, New York, 1993).
2. Stuardt L. Meyer, *Data analysis for scientists and engineers* (John Willey & Sons, Inc., New York, 1975).
3. D. C. Baird, *Experimentación*, 2ª ed. (Prentice-Hall Hispanoamericana S.A., México, 1991).
4. J. Higbie, “Uncertainty in the linear regression slope” *Am. J. Phys*. **59**, 184 (1991)
5. J. Orear, “Least squares when both variables have uncertainties”, *Am. J. Phys*. *ibid.*, **50**, 912 (1982).
6. “Simple method for fitting data when both variables have uncertainties” D. Barker and L.M. Diana Am. J. Phys. **42**, 224 (1974).
7. *“Linear least-squares fits with errors in both coordinates.” II: Comments on parameter variances* - B. Cameron Reed - Am. J. Phys., Vol. 60, No. 1, 1992
8. *Estadística* – M. Spiegel – McGraw Hill 2da. Ed. Bogotá 1997

Apéndice

Funciones de Excel – Incluidas en la planilla de cálculo de Excel *error_sg2k1.xls*, en el módulo Regression_sg, escrito en Visual Basic for Application se definieron las funciones que se describen a continuación y que fueron definidas en el texto de esta introducción. En dichos programas las dimensiones de los vectores son de 300, si se trabaja con vectores de dimensiones mayores, se deben modificar dichos programas. A estas funciones se accede a través del botón pegar funciones, sección: definidas por el usuario. Si se desea ver el código en Visual Basic, el menú de herramientas de Excel, elegir Macro: editor de Visual Basic. La mayoría de las funciones tiene incluida la opción *Mode* que se define a continuación:

mode=1	*No Weight*	$w_i = 1$
mode=2	*Weight*	$w_i = 1/\Delta Y_i^2$
mode=3	*Weight*	$w_i = 1/Abs(Y_i)$
mode=4	*Weight*	$w_i = 1/Y_i^2$
mode= 5	*Weight*	$w_i = abs(\Delta Y_i)$

$$W = \sum_i w_i \qquad \text{(A1)}$$

En general suponemos que los datos son resultados de *N* mediciones representados por las ternas ($X_i, Y_i, \Delta y_i$) las cuaternas ($X_i, Y_i, \Delta y_i, \Delta x_i$), donde los N datos X_i representan la variable independiente, representados genéricamente por el vector *Xdat*. Las errores asociados a la variable independiente, viene dados por los valores Δx_i , genéricamente representado por el vector Dx_dat. De modo análogo se definen N datos Y_i que representan la variable dependiente, representados genéricamente por el vector *Ydat*. Las errores asociados a esta variable dependiente, los designamos Δy_i , genéricamente los representamos por el vector DY_dat. Usando estas convenciones definimos las siguientes funciones:

a_Lfit(Xdat, Ydat, DYdat, mode) = Pendiente de la recta de regresión ec.(5):

$$a\_Lfit = a = \frac{W \cdot \left(\sum YX\right) - \left(\sum X\right) \cdot \left(\sum Y\right)}{W \cdot \left(\sum X^2\right) - \cdot \left(\sum X\right)^2} = \frac{Cov(YX)}{S_X^2} \qquad \text{(A2)}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *pendiente* de Excel.

b_Lfit(Xdat, Ydat, DYdat, mode) = Ordenada en el origen de la recta de regresión ec.(6):

$$b\_Lfit = b = < Y > - a \cdot < X > \qquad \text{(A3)}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *Intesección.eje* de Excel.

Da_Lfit(Xdat, Ydat, DYdat, mode) = Error en la pendiente de la recta de regresión ec.(5) y (29):

$$\Delta a_0 = Da\_Lfit = \sqrt{\frac{\chi_N^2}{(N-2)}} \cdot \frac{1}{S_X} = \frac{sd}{\sqrt{N} \cdot S_X} \tag{A5}$$

Db_Lfit(Xdat, Ydat, DYdat, mode) Error en la pendiente de la recta de regresión ec.(6) y (30):

$$\Delta b_0 = Db\_Lfit = b = \Delta a_0 \cdot < X^2 > \tag{A6}$$

Prom_pesado_n(Ydat, DYdat, nn, mode)

$$\Pr om\_pesado\_n(X) \equiv < X^n > = \frac{\sum_i w_i \cdot X_i^n}{W} \tag{A7}$$

Varianza(Ydat, DYdat, mode)

$$Varianza(Y) = \frac{N}{(N-1)} \cdot \frac{\sum_i w_i \cdot (Y_i - \bar{Y})^2}{W} \tag{A8}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *var* de Excel.

SYX_corr (Ydat, Yfit, DYdat, mode)

$$SYX\_corr = \sqrt{\left(\frac{n}{n-2}\right) \cdot \sum_i w_i \cdot \left(y_i - y(x_i)\right)^2 / W} = \sqrt{\left(\frac{N}{N-2}\right) \cdot \left(S_Y^2 - \frac{S_{YX}^2}{S_X^2}\right)} \tag{A9}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *error.tipico.YX* de Excel.

SYnXm (Ydat, Xdat, DYdat, nn, mm, mode)

$$SYnXm = \frac{\sum_i w_i \cdot \left(y_i^n \cdot x_i^m\right)}{W} \tag{A10}$$

Prom_pesado_n(Ydat, DYdat, nn, mode)

$$SYnXm = \frac{\sum_i w_i \cdot \left(y_i^n \cdot x_i^m\right)}{W} \tag{A11}$$

R_lin (Xdat, Ydat, DYdat, mode)

$$R\_lin = \equiv \frac{<YX> - <Y><X>}{S_X \cdot S_Y} = \frac{S_{YX}}{S_X \cdot S_Y}\,. \qquad \text{(A12)}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *Coeficiente.de.correl* de Excel.

regress_2(Ydat, Yfit, DYdat, mode)

$$regress\_2 = R^2 = 1 - \left(\frac{\sum w_i \cdot (Y_i - Y(X_i))^2}{W} \right) \Big/ \left(<Y^2> - <Y>^2 \right) \qquad \text{(A13)}$$

Cuando todos los datos tienen igual peso, esta función coincide con la función *Coeficiente.R2* de Excel.

Function Chi_un (Ydat, Yfit, DYdat, mode, Nparam)

$$Chi\_Nu = \left(\frac{N}{N - N\_param} \right) \cdot \frac{\sum w_i \cdot (Y_i - Y(X_i))^2}{W} \qquad \text{(A14)}$$

Chi2_tot (Ydat, Yfit, DYdat, mode)

$$Chi2\_tot = N \cdot \frac{\sum w_i \cdot (Y_i - Y(X_i))^2}{W} \qquad \text{(A15)}$$

Covar_YX (X_datos,Y_Datos, DY_dat, mode)

$$Cor\_YX \equiv Cov(Y,X) = \sum_{i=1}^{N} w_i \cdot \left(X_i - \bar{X} \right) \cdot \left(Y_i - \bar{Y} \right) \Big/ W = <YX> - <Y><X>\,, \qquad \text{(A16)}$$

SYnXm (Ydat, Xdat, DYdat, nn, mm, mode)

$$SYnXm = \sum_{i=1}^{N} w_i \cdot X_i^n \cdot Y_i^m \Big/ W = <Y^n X^m>\,, \qquad \text{(A17)}$$

Dy_lim_confianza(X_new, tp, Xdat, Ydat, DYdat, mode)

$$Dy\_\lim\_confianza(X, tp, X_i, Y_i, \Delta Y_i, Mode) = tp \cdot \frac{sd}{\sqrt{N-2}} \cdot \sqrt{1 + \frac{(X - X)^2}{N \cdot S_X^2}}\,, \qquad \text{(A18)}$$

dy_estima_media(X_new, tp,Chi_n,N, X_bar,Var(X))

$$dy\_estima\_media(X, tp, \chi^2_{N\,i}, N, \bar{X}, Var(X)) = tp \cdot \frac{\sqrt{\chi^2_N}}{\sqrt{N-2}} \cdot \sqrt{1 + \frac{(X - X)^2}{S_X^2}}\,, \qquad \text{(A19)}$$

dy_estimacion(X_new, tp,Chi_n,N, X_bar,Var(X))

$$dy\_estimacion(X, tp, \chi^2_{N\,i}, N, \overline{X}, Var(X)) = tp \cdot \sqrt{\frac{\chi^2_N}{(N-2)}} \cdot \sqrt{1 + N + \frac{(X-X)^2}{S_X^2}}\,, \quad \text{(A20)}$$

Slop_1d(Xdat, Ydat, W_dat)

$$Slop\_1d = a = \frac{W \cdot \left(\sum YX\right) - \left(\sum X\right) \cdot \left(\sum Y\right)}{W \cdot \left(\sum X^2\right) - \cdot \left(\sum X\right)^2} = \frac{Cov(YX)}{S_X^2} \qquad \text{(A21)}$$

Slop_2D(Xdat, Ydat, Wy_dat, Wx_dat)

$$Slop\_2d = a = \frac{W \cdot \left(\sum YX\right) - \left(\sum X\right) \cdot \left(\sum Y\right)}{W \cdot \left(\sum X^2\right) - \cdot \left(\sum X\right)^2} \qquad \text{(A22)}$$

La rutina que calcula la pendiente en esta caso, realiza 10 iteraciones, variando los pesos w_i, como las descriptas en el texto.